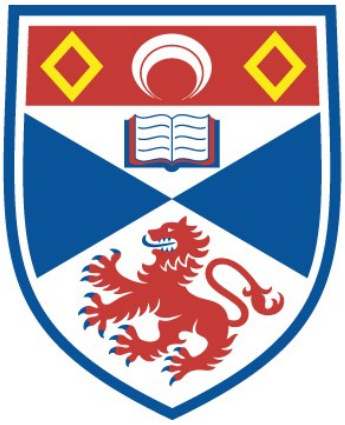
THE EFFECT OF SOCIAL STEREOTYPES ON EYEWITNESS BEHAVIOUR

Jason N. Frowley

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1994

THE EFFECT OF SOCIAL STEREOTYPES ON EYEWITNESS BEHAVIOUR

JASON N FROWLEY

Ph.D

I, Jason Frowley, hereby certify that this thesis, which is approximately 98,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 02.03.94 signature of candidate

I was admitted as a research student under Ordinance No. 12 in October, 1990 and as a candidate for the degree of Ph.D in October, 1991; the higher study for which this is a record was carried out in the University of St Andrews between 1990 and 1993.

date 02.03.94 signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date: 28 June, 1994 signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date 02.03.94 signature of candidate

ACKNOWLEDGEMENTS

Thanks to Amanda Barrie, Crispin Bonham-Carter, Dugald Bruce-Lockhart, Julian Crowe, Vincent Connelly, Oscar de Bruijn, Mary-Margaret Giannini, Athene Gordon, Hugh Greenway, Alice Hogge, Emma Mack, Malcolm MacLeod, Sophia Peck, Corinna Peniston-Bird, Mark Snyder, Will Sussman, Fergus Webster, all the departmental secretaries, my mother and father, and of course, Wellington.

CONTENTS

Abstract viii

Chapter 1 - Introduction 1

Overview 1

Social Factors in the Eyewitness Situation 1

The Social Stereotyping Literature 13

Summary 15

Chapter 2 - Stereotypes and Judgment: A Review of the Literature 17

Introduction 17

Social Functions of Stereotypes 18

Cognitive Functions of Stereotypes 26

Bodenhausen (1988) 27

Macrae (in press) - A Response to Bodenhausen 32

When the Hypotheses Hold 34

Judgments Made Outside the Juridic Situation 39

Conclusions 42

Chapter 3 - Does Size Matter? The Effect of Body-Build on the Judgment of an Ambiguous Action 45

Introduction 45

Literature Review 47

Pilot Study 57

Heider and Simmerl (1944) 59

Duncan (1976) 60

Experiment 1 64

Methodology 66

Results 74

Discussion 86

Chapter 4 - Building Offender Profiles: The Effect of Stereotypic Information on the Regeneration and Judgment of Facial Stimuli 93

Introduction 93

Literature Review 93

Experiment 2a 120

Methodology 121

Results 123

Discussion 127

Experiment 2b 130

Methodology 131

Results 131

Discussion 135

General Discussion 136

Chapter 5 - Every Picture Tells a Stereotype: The Effect of Stereotypic Information on the Regeneration and Judgment of Whole-Body Stimuli 139

Introduction 139

Literature Review 139

An Outline of Experiment 3 150

Experiment 3a 152

Methodology 153

Results 156

Discussion 160

Experiment 3b 162

Methodology 163

Results 164

Discussion 166

General Discussion 168

Chapter 6 - Stereotypes and Memory: A Review of the Literature 172

Introduction 172

Stereotypes Presented at Encoding 174

Stereotypes Presented at Retrieval 186

Summary and Conclusions 206

Social Factors in the Police Interview 207

Chapter 7 - Reconstruction of Things Past: Retrieval Schemata and the After-Effects of Stereotyping Information 214

Introduction 214

Experiment 4a 222

Methodology 223

Results 227

Discussion 234

Experiment 4b 237

Methodology 238

Results 240

Discussion 241

General Discussion 244

Chapter 8 - How the Reconstruction of Things Lasts: Delay, and the Effects of Stereotypic Information Presented at Retrieval 248

Experiment 5 248

Introduction 248

Methodology 249

Results 250

Discussion 257

Chapter 9 - Kiss and Tell: The Effect of Stereotypic Information on Memory for a Staged Incident 263

Introduction 263

An Outline of Experiment 6 264

Methodology 271

Results 275

Discussion 282

Summary and Conclusions 286

Chapter 10 - Interviewing Eyewitnesses: A Review of the Literature 289

The "Optimal Police Interview" 290

The Police Interview 292

Techniques to Aid Retrieval 301

Hypnosis 303

Contextual Reinstatement 309

Conversation Management 312

The Cognitive Interview 315

Conclusions 324

Chapter 11 - Cut on the Bias: Use of the Source-Monitoring Technique to Reduce Stereotypic Distortion in Recall 327

Introduction 327

Literature Review 329

Experiment 7 343

Methodology 343

Results 349

Discussion 366

Chapter 12 - Citing Your Sources: Use of the Source-Monitoring Technique in a Delayed-Test Condition 371

Experiment 8 371

Introduction 371

Methodology 375

Results 375

Discussion 392

Chapter 13 - Discussion 401

Introduction 401

Judgment 401

Memory and Interview Techniques 408

Towards a Model of the Effects of Stereotypes in the Eyewitness Situation 419

Directions for Future Research 428

Bibliography 432

Appendices 465

LIST OF FIGURES

Figure 3.1 Perceived body-size pairings x mean perceived intensity of actions 79

Figure 3.2 Attributions made x perceived body-size of targets 80

Figure 3.3 Actual body-size pairings x mean perceived intensity of actions 84

Figure 3.4 Attributions made x actual body-size of targets 85

Figure 4.1 Target face with subsequent regenerations, "assault" and "theft" conditions 125

Figure 4.2 Target face with subsequent regenerations, "2i" and "2ii" conditions 133

Figure 7.1 Lesbian vs heterosexual hits, two test procedures, heterosexual label 228

Figure 7.2 Lesbian vs heterosexual hits, two test procedures, lesbian label 229

Figure 7.3 Lesbian vs heterosexual false alarms, two test procedures, heterosexual label 230

Figure 7.4 Lesbian vs heterosexual false alarms, two test procedures, lesbian label 231

Figure 7.5 Neutral hits, two test procedures, two labels 233

Figure 7.6 Neutral false alarms, two test procedures, two labels 234

Figure 8.1 Lesbian vs heterosexual hits, two test times, two labels 252

Figure 8.2 Lesbian vs heterosexual false alarms, two test times, two labels 253

Figure 8.3 Neutral hits, two test times, two labels 255

Figure 8.4 Neutral false alarms, two test times, two labels 256

Figure 11.1 Visual, conversational and factual information x group x phase 351

Figure 11.2 Factual information recalled x information type x phase 352

Figure 11.3 Visual, conversational and factual errors made x group x phase 357

Figure 11.4 Factual errors made x information type x phase 358

Figure 12.1 Correct recall of factual information x subcategory x phase 377

Figure 12.2 Factual information recalled x information type x phase 378

Figure 12.3 Visual, conversational and factual errors x group x phase 382

Figure 12.4 Factual errors made x information type x phase 383

ABSTRACT

Eyewitness behaviour is a very important issue in social psychology. Recent years have seen a boom in research in this area; however, very little of this research has addressed the important and fundamental issues raised by the social factors that are involved when eyewitnessing takes place. The present thesis reports research which addresses the effect of social-stereotypic information upon a number of aspects of the eyewitness situation. Of particular interest is the effect of stereotypic information upon the judgment of aggressive and violent actions, and face and body stimuli; on memory for information associated with a dramatic staged incident, when presented either before or after target material; and its impact in an eyewitness interview situation. Finally, an interview technique is designed and tested which aims to reduce the negative impact of stereotypic information on eyewitness memory. Throughout, the impact of stereotypic information presented at encoding and at retrieval is contrasted and compared. The effect of delay between encoding and retrieval is also considered. It is concluded that stereotypic information may affect judgments of information relevant to the eyewitnessing situation whether presented before or after target material. Similarly, stereotypic information may affect memory when it is presented before target material. Little effect is however to be expected on memory when stereotypic information is presented at retrieval, although the exact effects may be found to vary with the nature of the target stimuli used. The source-monitoring interview, based on research by Marcia Johnson and her colleagues, appears a promising means of reducing the distorting effect of stereotypes on memory; particularly when a delay occurs between the encoding and retrieval of target material.

CHAPTER 1 - INTRODUCTION

Overview

It is argued in the present chapter that too little recent research in the eyewitness field has addressed the important issue of the impact of social factors. The present thesis represents an attempt to increase our knowledge of this important area by a preliminary investigation of the impact of stereotypic information upon judgment and memory factors in the eyewitness situation, and also the production of an interview technique which aims to reduce these effects. The thesis falls into three sections, each dealing with one of these areas. An introductory chapter to each section reviews the relevant literature in the area and its importance in the eyewitnessing situation. The aim of the present chapter then is briefly to outline the current state of the eyewitness literature, to identify the type of question to be considered later in the thesis, and to point up the need for further research into the importance of social factors to the eyewitness situation.

Social Factors in the Eyewitness Situation

Recent years have seen something of a research boom in the eyewitness memory field. This is understandable given the great social importance of the criminal justice system, and its heavy reliance (some, e.g., Woocher, 1977, would call it over-reliance) on eyewitness testimony. According to Sobell (1972,

cited in Geiselman, Fisher, MacKinnon and Holland, 1986,) incorrect eyewitness identifications may have been responsible for more miscarriages of justice than all other factors combined. A new research paradigm with implications for the area has also emerged in recent years - this is the social cognition paradigm. The aim of social cognitive research is "to push cognitive explanations...as far as possible" (Johnston and Hewstone, in press). The application of this research paradigm to the eyewitness situation has been responsible for a large part of the work of the last decade or so (see, e.g., Bodenhausen, 1988; Bodenhausen and Wyer, 1985; Macrae in press; Macrae and Shepherd, 1989a and b).

As Sporer (1982) has shown, the relatively recent flowering of the "psychology of testimony" is in many ways a renaissance of the "*Psychologie der Aussage*" that flourished at the beginning of the century. Although the problems of eyewitness testimony seem to have been known even to the historiographer Thucydides in the fifth century BC (Levine and Tapp, 1973), laboratory studies in the area were not conducted until the last years of the nineteenth century. It would probably be wrong however, as Sporer (1982) points out, to try to finger any one figure as the "father" of such research: more or less concurrently, Cattell studied "incidental memory" - aspects of which have relevance to the eyewitness situation; Bolton investigated the confidence - accuracy relationship; examining justice Hans Gross performed early "witness

tests" with the witnesses in his own criminal investigations (Cattell, 1895; Bolton, 1896; Gross 1894/1907; all cirted in Sporer, 1982). It is likely that more textbooks cite Alfred Binet than any other researcher as the first important figure in the "*Psychologie der Aussage*". Indeed it is interesting to note that two aspects of his work have since developed into central areas of research in the field: those of the child witness, and the suggestibility effect (Loftus, 1977; Schooler, Gerhard and Loftus, 1986; McCloskey and Zaragoza, 1985/87; Lindsay and Johnson, 1989; Wright, 1992). Binet's work is often cited today in support of the traditional view of the extreme suggestibility of children: However, as Cunningham (1988) points out, his work has never been translated into English, and citations of his conclusions take them largely at face value and from secondary sources. Binet was the first experimenter in the literature to employ what Cunningham terms a "multiobservational methodology": his aim, throughout several experiments, was to discover the relationship if any, between age, academic aptitude, and suggestibility, for male elementary schoolchildren aged 7 - 14 years. Using such methodology, Binet was the first to discover such now-familiar effects as qualitative differences between reports given spontaneously and those given in response to specific questions, the powerful effect of the form of a question upon the answer given, and the group conformity effect. It was a source of disappointment

to Binet that he was never really seen as a prophet in his native land; while research into eyewitness testimony had its origins in France, it was left to German researchers such as Louis William Stern to develop it. From an historical perspective, Cunningham remarks:

> "...*La Suggestibilité* marked a radical turning point in which the position of Bernheim and the Nancy school on the normalcy of suggestibility could finally be accepted. As such, it laid a cornerstone for the scientific study of influence and other aspects of social behavior (Apfelbaum, 1985). The misfortune of Binet's death in 1911, however, ablated the vision he had for the progress of social psychology (Binet, 1909); and the work's significance in the founding of modern experimental social psychology was never fully developed (Apfelbaum, 1985)" (Cunningham, 1988, p 275).

Stern appears to have been the second important figure in the history of eyewitness memory research. His work is reviewed in some detail by Sporer (1982), and is considered only briefly here. Stern (1902) was the first to stress the less-than-perfect functioning of the memory in the eyewitnessing situation: which led him to suggest that the court of law should be more cautious in its acceptance of eyewitness testimony. Stern is particularly important as the inventor of the staged incident (*Wirklichkeitversuch*) research paradigm as a means of studying memory for dramatic events which are in some ways similar to witnessed crimes. This paradigm had a burst of popularity at the beginning of the century, and has returned to favour in the last twenty or so years.

Following Stern's "call for an interdisciplinary approach to the problem of *'Aussagepsychologie'*", (Sporer, 1982, p 328,) eyewitness memory became quite a fashionable research area: spreading as far afield as Russia, India, and Chile. Researchers during this period appear to have been keen to stress the practical implications of their work: something which has not always been the case in more recent years (cf. Yuille and Cutshall, 1986). The reaction from the legal profession appears to have been generally positive:

> "Their criticisms were to the point but constructive, emphasizing Stern's (1902) postulate of 'closeness to life' ('Lebensnähe') as a principle for experimentation and thus anticipating many arguments about the 'external validity' of laboratory experiments as we would call it today (Loh, 1981; Monahan and Loftus, 1982)" (Sporer, 1982, p 330).

While progress in the eyewitness testimony field has never come to a complete halt - indeed, some interesting work was published in mid-century, see, e.g., Rourke, 1957 - the dominance of the Behaviourist approach during these years resulted in an inevitable reduction of pace.

With the relatively recent increase in importance of cognitive psychology, the new area of social cognition has arisen; giving researchers in the area new tools with which to address eyewitness testimony issues. In 1983, McCloskey and Egeth published a paper entitled "Eyewitness identification: What can a psychologist tell a jury?"; implicitly answering their question

"not much". As noted above, however, recent years have witnessed an increase in research, and in 1991, Deffenbacher was able "to suggest that a psychologist can now tell police quite a bit" (p 377). While it is not the purpose of the present chapter to review the recent literature in its entirety, (Deffenbacher, 1991, presents perhaps the best recent review,) it is possible to mention briefly the areas in which most work has been done. Conspicuous failures of recent research have been in attempts to identify individual factors predictive of an eyewitness' ability to provide useful and accurate testimony; and to develop procedures by which to improve the encoding of a witnessed event (e.g., Deffenbacher, Brown and Sturgill, 1978; Malpass, 1981). For face recognition, a number of studies (e.g., Klatzky, Martin and Kane, 1986; Shapiro and Penrod, 1986) have shown a depth-of-processing effect at encoding: faces for which more cognitively complex processing occurs later show a memory advantage. However, short-term training in memorial strategies does not appear to be helpful (Woodhead, Baddeley and Simmonds, 1979). The storage of eyewitness memories has also received attention: e.g., the degree of forgetting with delay interval was studied by Deffenbacher (1986); and the nature of retroactive interference during misinformation studies was studied by McCloskey and Zaragoza (1985) and by Lindsay and Johnson (1989). Much more attention has been paid to issues associated with the retrieval of eyewitness memories, and specifically the development

of techniques designed to improve retrieval. Techniques that have been developed include hypnosis, conversation management, and the cognitive interview, both basic and enhanced. This literature is reviewed in some detail in Chapter 11. Additionally, techniques have been developed to aid police in cueing eyewitness memory: e.g., the use of face composite systems, such as Photofit and Identi-kit. A good deal of recent research has addressed the usefulness or otherwise of these systems (e.g., Davies, 1983; Davies, Ellis and Shepherd, 1978); findings tending to be somewhat negative (e.g., Christie and Ellis, 1981, found that Photofits were less accurate than witnesses' verbal descriptions of faces - see Chapter 4 for a fuller discussion of literature pertaining to this topic). A great deal of research has addressed factors involved with the identification parade, and how the fairness of such parades can be increased (see, e.g., Cutler and Penrod, 1988; Cutler, Penrod, O'Rourke and Martens, 1986; and Malpass and Devine, 1983).

Research has also addressed the role of what Wells (1978) terms "estimator variables" (i.e., those over which the criminal justice system has no control). Examples of this are studies of event characteristics (e.g., the length of a witness' exposure to a criminal - Ellis, Davies and Shepherd, 1977; and the effect of "weapon focus" - Loftus, Loftus and Messo, 1987). The individual characteristics of suspects has also received attention: e.g., the role of the suspect's

sex, race, and disguise (Shapiro and Penrod, 1986; Cutler, Penrod and Martens, 1987), as have the characteristics of witnesses. For example, Shapiro and Penrod appear to have established that race has no effect upon eyewitness accuracy; while as far as sex is concerned, females may have slightly superior memory for faces than males (see, e.g., Clifford and Bull, 1978, for a dated but detailed discussion of this). The age of eyewitnesses however appears to be of great importance (Chance and Goldstein, 1984; Flin, 1980). Much research (Bothwell, Deffenbacher and Brigham, 1987; Cutler and Penrod, 1989) has addressed the old issue of the correlation between an eyewitness' confidence and his/her accuracy - which appears to be positive but small - and also the correlation between a witness' accuracy of prior verbal description of a suspect and his/her subsequent identification accuracy - which does not appear to be significant.

Despite the apparent exhaustiveness of this list, however, as Tollestrup, Turtle and Yuille (1993) point out, recent research has tended to consider certain topics in far greater depth than others. There has been a particularly pronounced emphasis upon such areas as interviewing practices, identification procedures, the relationship between an eyewitness' confidence and her/her accuracy, racial issues, and the eyewitness abilities of young children. However, other areas appear to have been unjustly neglected. There has, for instance, been a tendency to

concentrate upon the negative rather than the positive aspects of eyewitness memory (Yuille and Cutshall, 1986), and upon memory for facial rather than body features (MacLeod, Frowley and Shepherd, in press). The above-mentioned use of the social cognition research paradigm, while carrying with it many undoubted advantages, has also tended to foster a greater reliance by eyewitness memory researchers on the *cognitive* rather than the *social*. Much research utilises the implicit model of the eyewitness as "black box", conceptualizing the ultimate aim of research in the area as detailing the cognitive apparatus contained within that box.

From an objective point of view, however, it might seem strange that more attention has not been devoted to the role of social information in the eyewitness situation. After all, eyewitnessing always takes place in a social situation, as does an eyewitness' recall of the event (whether among friends and relatives, or during a police interview). Social information might therefore be expected to have powerful implications both for eyewitness memory and for other aspects of a crime situation.

An example might help to elucidate this point. In the novel *Intruder in the Dust* by the Southern writer William Faulkner, a black man, Lucas MacCalsin, is accused of stabbing a white man in the back. This accusation is made on the basis of little more evidence than that Lucas *is* black, and this was considered more or less normal behaviour for blacks at

that time and in that part of the world. Lucas, however, will not even answer the charges that are brought against him. Proud of the white blood that is in his veins, he refuses to act according to the prevailing social stereotype of black people. A mob of angry "cognitive misers" (Macrae, in press) is therefore organised to go out and lynch him. This is, of course, just a novel: but, given the social situation prevalent in that part of the world (similar scenarios could be derived from any number of other sources) it is probably not a very far-fetched one.

Similar social stereotypes are applied particularly to the eyewitness situation by Boon and Davies (1987b); a study which demonstrates quite clearly what is meant by stereotypic information causing memory distortions. Part of the Boon and Davies study was based upon an earlier study by Allport and Postman (1945): a study which utilised the then-current racial stereotype of blacks as carriers of cut-throat razors. Boon and Davies wanted to assess the influence of an updated version of this stereotype (blacks as knife-carriers) on the reporting of a story. They recounted to one subject the story of a subway incident between a white person and a black person, and then used a chain-report technique to generate a final version of the story. In the original version it was specified that a knife was pulled, but not who pulled it. The result of interest was to see who was said to have pulled the knife at the end of the chain. They found that in no case was the white said to have been the

knife-carrier. The third experiment reported in this paper shows a similar kind of effect: Again, a social variable, this time attitude, was found to have an influence in the eyewitness situation. In this experiment, subjects were shown photographs of a mounted policeman attempting to break up a miners' strike, after having their subjective attitude towards the strike tested. Finally, they were shown a set of line-drawings based on the photograph, and asked to select from these the drawing which in their opinion best represented the photograph. The line-drawings differed only in how threatening a posture the policeman was adopting. Subjects who were more in favour of the strike tended to select drawings in which the policeman appeared more threatening, and vice versa.

The Boon and Davies study (reviewed in more detail in Chapter 10) is an elegant demonstration of how important social factors - attitudes or stereotypes - can be in the eyewitnessing situation. Similar points have been made by Lindsay and Johnson (1989), and by Shepherd, Ellis, McMurran and Davies (1978). Nevertheless, social factors have received relatively little attention in the eyewitness literature.

Only relatively recently has a literature even begun to develop concerning these more fundamentally social psychological aspects of the eyewitness situation, from the actual witnessing of the crime, through the police interview, to appearance in court. The influence of such a perspective - viewing the witness

as a human being intricated in a series of more or less complex and unpleasant social situations - can be seen in a number of disparate areas of recent research. Miller, Turnbull and MacFarland (1990) investigated the common phenomenon of counterfactual thinking ("thinking about what might have been"), and the factors which influence the instance of such behaviour: a research area with obvious applicability to the crime situation, both for the eyewitness, and, more importantly, the victim. Until relatively recently, the police interview also suffered from failing to identify itself as an ultimately social situation, in which factors such as dominance-submissiveness and poor conversational technique played profound roles (Adams, 1985). Recent advances in research into the police interview have however tended to emphasise the importance of conceptualising it in social terms: hence the birth of such strategies as the enhanced cognitive interview (Fisher *et al*, 1987; Geiselman and Fisher, 1988) and conversation management (E Shepherd, 1985/87; George 1992). The conceptualisation of the criminal trial as an intimidating social situation has meant that children may now present evidence via video link to the courtroom (Flin and Tarrant, 1989). Finally, Bodenhausen and Kramer (in press) present a study of the effects of stereotypic information in the criminal justice system, addressing such questions as whether social stereotypes can affect the veracity of eyewitness reports, and whether jurors' own prejudices

lead to differential learning of trial testimony and differential recall of this during their final deliberations. Nevertheless, despite recent advances, such studies are certainly in the minority. The present thesis attempts to go some way towards redressing this balance, by assessing the influence of the social phenomenon of stereotyping on several aspects of the eyewitnessing situation; in many cases, aspects which appear never to have received such attention in the past.

The Social Stereotyping Literature

There exists an immensely complex literature in the stereotyping field. An attempt to apply all of the extant research paradigms and perspectives (from psychodynamic models of stereotype formation, to social identity theory, to studies of illusory correlation) to the eyewitness situation would inevitably cause confusion and blur interpretation of the results achieved.

A simplifying approach is therefore adopted here. It is argued in Chapters 2 and 6 that two broad approaches to the stereotyping literature can be identified; these are discriminable essentially by the type of explanation of phenomena which they offer; explanations which can be more or less "social" or "cognitive" in nature. The social psychological approach has often dealt with issues which are not directly relevant in the present instance: e.g., many social psychological studies are essentially

descriptive, detailing as they do such phenomena as the contents of one group's stereotypes of another group (e.g., Katz and Braly, 1933; Gilbert, 1951; Karlins, Coffman and Walters, 1969). More relevant to the present situation are studies of the inter-relationship between social stereotypes and target information of one sort or another (e.g., ambiguous actions - Duncan, 1976, Sagar and Schofield, 1978; biographical information - Snyder and Uranowitz, 1978c, Bellezza and Bower, 1981; or faces - Bartlett, 1932). The literature has considered the effects of stereotypes upon both judgment of such target material (e.g., Bodenhausen and Lichtenstein, 1987; Bodenhausen and Wyer, 1985) and memory for this material (e.g., Cohen, 1977; Rothbart, Evans and Fulero, 1979). More ambitious studies have attempted to account for the wide-ranging results that have been produced both for judgment (e.g., Bodenhausen, 1988; Macrae, in press), and memory for target material (e.g., Hastie, Park and Weber, 1984; Stangor and Ruble, 1989; Wyer and Gordon, 1982/84).

The present thesis broadly follows this general conceptualisation of the previous research. Each of these three areas, (i.e., judgment, memory, and the construction of models to account for the findings,) is considered in some detail, both in a review chapter and empirical research chapters. Chapters 2 and 3 deal exclusively with judgment phenomena, although these are also dealt with less exclusively in Chapters 4 and 5; Chapters 7 - 9 deal with memory phenomena;

and Chapters 6 and 13 with the construction of models of the effects of stereotypic information upon memory for target materials. Additionally, Chapters 10 - 12 deal with an area of particular relevance to the eyewitness situation - that of the police interview - and how this can be used to reduce the impact of stereotypes upon memory for a witnessed crime.

The approach adopted in Chapters 6 and 13 is predominantly a social cognitive one. This is not to deny the basic importance of the social approach - indeed, as outlined above, stereotypes are conceived here as essentially social phenomena - but reflects the belief that cognitive models are a valid approach to this social psychological field. The construction and testing of models of stereotypic effects appears the optimum means of refining our conceptualisation of this highly important field.

Summary

Despite a recent boom in eyewitness research, little work has considered the importance of social factors in the eyewitness situation. The present thesis attempts a preliminary outline of the ways in which the social phenomena of stereotypes may impact upon this situation. The present state of research into stereotyping is highly complex, but it is possible to identify three more or less discrete foci and two more or less discrete research paradigms. The foci of particular interest here are the effects of stereotypes on judgment and on memory, and the

construction of models of these effects. The research paradigms are the social and the social-cognitive.

The present thesis consists essentially of three sections. The first section considers the impact of stereotypes on judgments in the eyewitness situation, and the second considers the impact of stereotypes upon eyewitness memory. Both of these sections also consider implications for the construction of models of such effects. The third section concentrates upon the police interview, and means of reducing the impact of stereotypes in the eyewitness situation. An introductory chapter to each section outlines relevant research in the area, devoting attention to both social and social-cognitive paradigms. A general discussion in Chapter 13 attempts to summarise and explain the results in terms of social cognitive models of stereotype action.

CHAPTER 2 - STEREOTYPES AND JUDGMENT: A REVIEW OF THE LITERATURE

Introduction

The present chapter attempts briefly to review some of the psychological literature on stereotypes and judgment relevant to the eyewitnessing situation. Two approaches to this question are discriminable: broadly these are the social (which dominated most earlier research on the topic) and the social-cognitive (which has been the dominant research paradigm in about the last two decades). While much research has been of a generally descriptive nature, (see the next sub-section,) the literature has dealt with a number of varied aspects of the stereotyping question. The role of the "authoritarian personality" in the formation of stereotypes was investigated by Adorno, Frenkel-Brunswik, Levinson and Sanford (1950); Sinha and Upahyaya (1960) studied the role of world events in changing stereotypes of ethnic outgroups; Snyder and Uranowitz (1978c) looked at the way in which stereotypes might provide retrieval schemata for previously-presented biographical information; Bodenhausen and Lichtenstein (1987) looked at the way in which stereotypes could mediate juridic judgments; Levine (cited in Tajfel, 1981,) looked at the role of economic factors in the creation of group stereotypes. The present thesis looks at only a narrow area of the stereotyping question; that is, the aspects which are of most relevance to the eyewinessing situation. Of particular importance are the distorting effects of

stereotypes on memory (Chapters 9 and 10), the use of stereotypes as retrieval schemata (Chapters 7 and 8), and their role in the formation of social judgments. This final area is the specific area of interest in the first part of the thesis. The first two sub-sections in the present chapter consider, respectively, the literature concerned with the social and the cognitive functions of stereotypes. While the thrust of this chapter is the consideration of the effects of stereotypes on social *judgments*, much of the research into the functions of stereotypes has considered their effects upon social *memory*. Therefore, it is necessary for the present chapter to make some mention of this literature, also.

Social Functions of Stereotypes

The greater part of early research into stereotyping behaviour was descriptive in nature, attempting to define the contents of particular groups' stereotypes of other groups (see, e.g., Allen and Wilder, 1975; Billig and Tajfel, 1973; Gilbert, 1951; Karlins, Coffman and Walters, 1969; Katz and Braly, 1933; Tajfel, Billig, Bundy and Flament, 1971; Taylor and Jaggi, 1974). These studies demonstrate stereotypic bias in terms of how members of particular social groups are perceived, and the way in which their behaviour or psychological characteristics are judged. More recent work has concentrated less on the contents of stereotypes as on providing explanations of how and why they arise. Tajfel (1981) characterises this work

as favouring one of two different types of explanation: the social or the cognitive. As Ashmore and Del Boca (1981) point out, the "social" orientation may further be broken down into "sociocultural" and "psychodynamic" approaches: however, for the sake of simplicity, this scheme is not adopted in this review. The present sub-section considers social explanations as a single unified approach.

This social approach can be characterised as being concerned with questions of a descriptive sort, such as "what are the contents of ingroup *x*'s stereotype of outgroup *y*?" (see, e.g., the list presented above) and more general theoretical questions of the sort "how do stereotypes affect perceptions of outgroup members?" (e.g., Jussim, Coleman and Lerch, 1987), and "what social needs does the act of stereotyping fulfill?" (e.g., Testa, Kinder and Ironson, 1987). The cognitive approach meanwhile has tended to consider questions of the sort "what mental processes give rise to stereotypes?" (e.g., Morgan, 1989,), and "how is stereotypic information organised in memory?" (e.g., Hastie, 1981). As Tajfel (1981) remarks, we can afford to ignore neither approach. We will never have a complete picture of the functions of stereotypes without considering both the cognitive and social aspects of the question. Indeed, it has been argued that the cognitive approach may be meaningless on its own, as the social circumstances which determine stereotypes is in constant flux (Haslam, 1985). An

attempt to produce a coherent theory of the "social functions of stereotypes" must, Tajfel argues, "bring together what is known from social psychology, social history, social anthropology and common sense" (Tajfel, 1981, p 160). Three such social functions of stereotypes can be identified, and examples are cited from the literature in support of these.

Social stereotypes can be used as part of an attempt to understand large, complex, and often distressing social events. A colourful example of the use of stereotypes in this fashion is the "explanation" given for the English plague of the seventeenth century - this was blamed on "untrustworthy" Scots (for reasons of their own) poisoning the wells of Newcastle (Thomas, 1973, cited in Tajfel, 1981, p 161). The outbreak of the Austrian civil war in 1934 was attributed to destructive Bolshevik tendencies; resulting in the hanging or internment of members of the Social Democratic party (Scheithauer, Woratschek and Tscherne, 1983).

Second, stereotypes can be used to justify actions (whether planned for the future or committed in the past) against outgroups. Tajfel illustrates this principle by reference to the work of Kiernan (1972), who finds that European conquest of foreign countries was justified by a stereotype of other races as in need of "advancement" through the good offices of the Europeans; and the belief of the English in India that "*en agissant avec des brutes il faut être brutal*". Hitler's Final Solution is perhaps the most virulent

historical example of a stereotype being used to justify negative behaviour towards an outgroup (Shirer, 1959).

The final social function of stereotypes is positively to discriminate the ingroup from outgroups. This will be particularly important at a time when such differences are seen to be being eroded, or when further justification is required for negative action (social, cultural, or political) against an outgroup. This is certainly the most extensively researched of the social functions of stereotypes. Tajfel (1957/59/69) developed a theory which holds that the effect of categorisation is the cognitive exaggeration of differences *between* categories, and the minimisation of differences *within* categories. Such an effect holds not only for physical stimuli such as lines (Tajfel and Wilkes, 1963,) and nonsense stimuli (Campbell, 1956), but also for *social* stimuli. Secord, Bevan and Katz (1956), for example, found that highly prejudiced white subjects tended to perceive the skin colour of blacks as darker than in fact it was: a fairly clear example of the exaggeration of between-category differences. Tajfel, Billig, Bundy and Flament (1971) asked: "Can the very act of social categorisation, as far as it can be identified and isolated from other variables, lead - under certain conditions - to intergroup behaviour which discriminates against the outgroup and favours the ingroup?" (Tajfel *et al*, 1971, p 151). Tajfel attempted to answer this question through the

construction of "minimal groups". His work (e.g., Billig and Tajfel, 1973; Tajfel and Billig, 1974) repeatedly demonstrated that the construction of groups along even such artificial axes as preference for the paintings of Klee or Kandinsky, or even the outcome of a coin-toss, is sufficient to cause ingroup bias. This effect is increased with the introduction into the social situation of further variables which are normally associated with groups (Rabbie and Horwitz, 1969).

Early explanations of these results centred around what Tajfel (1969/70) termed the "generic norm" of ingroup favouritism. For a number of reasons however (see Billig, 1976), this conceptualisation does not appear tenable. Later explanations (e.g., Tajfel, 1972; Turner, 1975) have centred around social identity theory. Every individual, the theory holds, tries to form and maintain a positive self-concept. This has two aspects: the personal and the social identity. In different situations, one or the other of these aspects of the self-concept will become more or less salient. The social identity of a group is always achieved with reference to some outgroup. A social category marks the ingroup off from outgroups; and, furthermore, allows for explicit comparison of the groups along one or more dimensions. Denigration of the outgroup causes elevation of the ingroup, and consequent raising of social esteem. This process can work only if subjects actively engage in a social comparison process, and it is a strength of this

approach that it explains the counter-intuitive finding that in a social categorisation situation, subjects often act not only to maximise ingroup benefits, but also intergroup differences (Turner, 1975). Such phenomena as ingroup favouritism are therefore seen as consequences of attempts to achieve or maintain a positive social identity. Such effects may be explicable in terms of norms: when people are assigned to in- and outgroups, the concept of *teams*, and therefore the competitive norm, are evoked (Wetherell, 1981, personal communication). While the theory has not been without its critics (see Aschenbrenner and Schaeffer, 1980, and Brown, Tajfel and Turner, 1980), its major supposition - that of a link between intergroup discrimination and self-esteem - has been experimentally demonstrated on at least two occasions (Oakes and Turner, 1980; Lemyre and Smith, 1985).

Social identity theory appears therefore to provide, as Brown (1988) remarks, a plausible account of people's readiness to favour even the most minimal of ingroups. Neither is its applicability limited to the somewhat artificial situations which have tended to be used in the laboratory: it is also attractive insofar as Tajfel (1982) and Brown (1984) have shown it can also account for other, real-world phenomena. Indeed, it has been shown even outside the laboratory that the degree of matching of in- to outgroup attitudes may be negatively correlated with intergroup animosity (Thompson, 1988). Similar effects have also been

demonstrated by Taylor and Jaggi (1974) and by Hewstone and Ward (1985); and may be maintained by illusory correlation (Chapman, 1967; Hamilton, 1981): A person's social world may consist of members of in- and outgroups performing positive and negative actions. Ingroup members may be encountered twice as often as outgroup members, and positive actions twice as often as negative actions. Where two relatively rare occurrences (an outgroup member and a negative action) are encountered together, this event will be particularly memorable and their instance of co-occurrence will be over-estimated.

Another possible explanation of ingroup bias is that proposed by Dion (1979), using an analogy to the Gestalt in perceptual psychology. Categorisation is held to divide the social environment into mutually exclusive categories: ingroups and outgroups. Since the categories do not overlap, subjects expect to encounter social information which differentiates between them. Also, they will tend to act so as to accentuate these perceived differences. Intergroup discrimination, according to this view, is one means of maintaining cognitive differentiation between groups.

Views such as these might be taken to imply a social function or significance in social memory and indeed social "forgetting". Differential memory of actions of in- and outgroup members would tend to enhance the subjective social worth of the ingroup and denigrate that of the outgroup, so long as positive instances of

ingroup behaviour are better remembered than negative instances, and vice versa. Also, such a process might maximise the perceived difference *between* social groups. Such effects would be particularly relevant to the eyewitness situation.

Howard and Rothbart (1980) found that the simple categorisation of persons into in- and outgroups is sufficient to bias their recall of information about the groups. This bias is found to favour the ingroup: more "positive" descriptors are recalled that applied to the ingroup; and more "negative" descriptors that applied to the outgroup. Wilder (1981) reports an experiment in which subjects were better able to recall information relating to ingroup similarities and outgroup dissimilarities (relative to themselves) than vice versa. On a recognition task, they made more false identifications of characteristics that implied ingroup similarity and outgroup dissimilarity (to themselves) than vice versa. Such an effect may hold even for material with less external validity, e.g., Tsujimoto (1978). The more salient is category membership, the more powerful the effects are likely to be (Cantor and Mischel, 1977; Haslam, 1985). The effects of stereotypes on memory are discussed more fully in Chapter 6.

Despite this wealth of research, however, as Tajfel (1979) remarks, the purely "social" approach to stereotyping has yet to address some basic questions concerned with the occurrence of stereotypes.

Questions which remain unanswered are of the sort:

> "Why, when and how is a social categorisation salient or not salient? What kind of shared constructions of social reality, mediated through social categorisations, lead to a social climate in which large masses of people feel that they are in long-term conflict with other large masses? What, for example, are the *psychological* transitions from a stable to an unstable social system?" (Tajfel, 1979, p 188)

Equally, the social approach has been unable to answer the question of *how* social stereotypes operate. For an answer to this question, it is necessary to look at social cognitive research.

Cognitive Functions of Stereotypes

During the last two decades, a new approach to the study of social phenomena has arisen: this is the study of social cognition. This involves essentially the application to the social sphere of paradigms drawn from experimental cognitive psychology. Bodenhausen and Kramer (in press) define this field as having "the ambitious objective of conceptualising human social behaviour in terms of underlying information-processing dynamics". While it may therefore appear simply an application of cognitive psychology, the social element of social cognition has been defended by Leyens and Codol (1988) on the grounds that: 1) it has a social origin, being created or reinforced through social interaction, 2) it has a social object, and 3) it is socially shared, being common to different members in a given society or group.

Ultimately, social cognition provides a second, useful approach to issues which remain social psychological ones. The approach lends itself to the construction of models of what exactly takes place in the mind of the perceiver during a social situation. The construction of such models allows powerful predictions to be drawn regarding such important social events as eyewitnessing. It may prove possible for future researchers to construct a detailed model of the effects of stereotypic information upon eyewitness performance: both for memory and judgment factors. For this to occur, however, it will be necessary to draw upon a formulation of how stereotypic information can affect eyewitness memory and judgment. To this end, the remainder of the present chapter considers social cognitive approaches to the stereotypes and judgment literature.

Bodenhausen (1988)

Bodenhausen gives three models of the way in which stereotypes might affect judgments made about a member of a stereotyped group. These are reviewed below.

1. The interpretive hypothesis: Central to this hypothesis is the fact that a single piece of behaviour may be amenable to multiple interpretations (by different perceivers, or by the same perceiver at different times or in different states). Prior expectations or stereotypical beliefs can affect the interpretation of any given piece of behaviour; thus affecting subsequent judgments based upon that

behaviour. This hypothesis receives support from a number of studies; perhaps most importantly from studies of the different interpretations of the behaviour of black and white actors by Duncan (1976) and by Sagar and Schofield (1980) - for further discussion of these experiments, see Chapter 3. Support is also gained from an experiment reported by Bodenhausen (1988) himself - see below for a discussion of this.

2. Selective processing hypothesis: According to this model, the mere activation of a stereotype in a social situation is sufficent to disrupt the (otherwise relatively smooth) flow of social information processing, in such a way that perceivers find it easier to process stereotype-confirming than stereotype-disconfirming evidence. Confirming evidence is processed more elaborately and hence will prove easier to recall (Bradshaw and Anderson, 1982; Hastie, 1980). This recall advantage for confirming evidence means that judgments are also biassed. This hypothesis receives support from Bodenhausen and Wyer (1985), who investigated the effects of racial stereotypes on causal attributions, sanctioning judgments, and memory for criminal events. They found that when a criminal's race was stereotypically consistent with the crime of which he was accused (e.g., an Hispanic charged with assault), subjects made more dispositional attributions and recommended harsher punishments than when this was not the case

(e.g., a White Anglo-Saxon Protestant charged with assault).

Some methodological shortcomings of this study were unearthed by Macrae and Shepherd (1989a); but their conceptual replication of the study found a generally similar pattern of results. Bodenhausen and Lichtenstein (1987) had subjects judge the guilt or innocence and the aggressiveness of a defendant after receiving both incriminating and exculpating evidence. Prior to the task, subjects were told that they would be making only the complex guilt/innocence judgment or the simple trait judgment; however, all subjects actually made both judgments. When subjects thought they had only a simple judgment to make, ethnicity had little effect; however, when they thought they had a complex judgment to make, ethnicity had a similar effect to that discovered by Bodenhausen and Wyer (1985). Hence, stereotypic information appeared to be used only when information-processing demands were high.

Additionally, the studies of the illusory correlation phenomenon reported by Hamilton and his colleagues (e.g., Hamilton, 1981a; Hamilton and Gifford, 1976; Hamilton and Rose, 1980, and see Chapter 6) can be seen as demonstrating a kind of selective processing effect; providing this hypothesis with one more line of support.

3. The heuristic hypothesis: According to this hypothesis, stereotype activation may, in and of itself, be sufficient to cause biassed judgments. The very activation of a social stereotype provides the perceiver with a range of (more or less accurate or useful) information (e.g., Cantor and Mischel, 1977), upon which later judgments may depend, with little regard for the external validity of this information. Two experiments using stimulus materials based on rape cases provide evidence for this hypothesis. Ugwuegbu (1979) assessed how jurors made judgments about same- and different-race rapists, finding that other-race rapists were judged more harshly - in the absence of qualitatively different evidence against them. This effect appears to be mediated by the simple activation of a negative racial stereotype. And Jacobson (1981) found that, in judging rape cases, both male and female jurors gave greater credence to the alibi of an attractive than an unattractive defendant, found an attractive defendant guilty less often, and treated him more leniently when found guilty, presumably on the basis of a fairly primitive "beautiful-is-good" stereotype (an explanation which may also be cast in terms of implicit personality theory - Schneider, 1973).

Macrae (in press) points out that there is a second possible variant of the heuristic hypothesis, according to which the social perceiver may actually store precomputed social judgments in memory. The activation of a stereotype would therefore simply

entail the activation of a social judgment about the target, with no need for the "cognitive miser" to indulge in further laborious information processing. Empirical support can be found for both variants of this hypothesis: Bodenhausen, 1990; Bodenhausen and Wyer, 1985; Kahneman, Slovic and Tversky, 1982; Macrae and Shepherd, 1989a; Nisbett and Ross, 1980; Petzold, 1986; Sherman and Corty, 1984.

Empirical tests: Bodenhausen (1988) tests these three hypotheses in a pair of experiments, with particular emphasis upon legal judgments. As in his previous work, he had subjects act as mock jurors in cases of alleged assault in which the defendant was described as stereotypically linked or not linked with the crime (i.e., Hispanic or ethnically nondescript). He points out that both the biassed interpretation and selective processing hypotheses require the stereotype to be activated prior to receipt of other information. For the heuristic hypothesis, however, it does not matter when the stereotype is activated. Bodenhausen was therefore able to manipulate the time of stereotype activation to discriminate between the heuristic and other hypotheses. He found that stereotypic information had little effect upon juridic judgments (or recall of case information) when presented after receipt of other evidence. However, when it was activated *before* receipt of other evidence, stereotypic information had a powerful impact upon both juridic judgments and recall of case evidence.

Additionally, it was found that activation of a stereotype prior to receipt of incident-related information had no effect on ratings of probative implications of the presented evidence (e.g., "How unfavourable for the defendant would it be to learn that he left the bar ten minutes before the assault occurred?"). This refuted the selective processing hypothesis; hence Bodenhausen's study tended to support the biassed interpretation model.

Macrae (in press) - A Response to Bodenhausen

Macrae presents what is in part a review and critique of Bodenhausen's work, during which he makes the point that Bodenhausen's approach may be to some extent invalid. Specifically, Bodenhausen may be mistaken in testing both recall of case material, and judgments based upon that material, in the belief that the two are correlated. There may be grounds for believing that memory and judgment are wholly independent processes, for which different items of information are employed. The valency (positive or negative) of a judgment made on the basis of a stereotype need not, e.g., be related to the valency of the material which can be remembered with regard to that stereotype:

> "A central feature of Bodenhausen's conception of stereotypical effects on decision-making is that jurors' judgments and recall are correlated. Within the context of a 'story-telling' model, stereotype activation affects the mental representation formed of the case evidence. For example, when the crime is stereotypically-consistent with the defendant's category membership the mental representation of the case is likely to be biased towards the organisation of prosecution (i.e., incriminating) evidence.

> In recall, the accessibility of this evidence will increase the likelihood of jurors returning a guilty verdict (i.e., judgment/recall correlation). Thus, jurors' verdicts are related to their recall of case evidence."
> (Macrae, in press.)

This methodological weakness can also be seen in a review paper presented by Higgins and Bargh (1987).

In their study of this relationship, Hastie and Park (1986) found that a positive correlation is likely to obtain only in "memory-based" tasks: tasks which require social information to be stored in memory before judgments are made on the basis of that information. Such tasks may rarely be encountered outside the psychology laboratory. For "on-line" tasks, in which judgments are made as information is encountered, (as in the majority of real-world tasks) there will likely be little or no correlation between recall and judgment.

Jury decision-making appears to be unusual for a real-world task in that it contains both "memory-based" and "on-line" components (Hastie and Park, 1986); hence as far as this very particular case is concerned, Bodenhausen's conclusions appear likely to be correct. Other real-world cases however (e.g., the eyewitnessing situation) appear less likely to be structured in quite this way.

This is not of course to hold that Bodenhausen's conclusions are necessarily incorrect where it comes to the eyewitness situation, but simply that we should be careful in applying them in a hard-and-fast manner. Researchers must consider areas other than the very

specific one of jury decision-making. This implies that consideration should be given to the heuristic hypothesis, abandoned by Bodenhausen. Hence, the possible effects of time of stereotype activation must be addressed.

When the Hypotheses Hold

The present chapter has so far considered only the ways in which stereotypes might affect judgment of social information. No consideration has yet been given to the circumstances under which these effects are likely to hold. Bodenhausen and Wyer (1985) provide three fairly general hypotheses with regard to this:

1. the (confusingly-named) heuristic hypothesis, which in this case holds that people will look for other (situational or dispositional) explanations of a target's behaviour only when a stereotype-based explanation is unavailable. If a crime or other transgression can be interpreted in terms of a stereotype, it will be punished accordingly, regardless of what other information is to hand.
2. the default hypothesis, which holds that stereotypes are used only as a last resort in judgment tasks. The influence of stereotypes on judgment will be eliminated if other information is to hand.
3. the integration hypothesis, which holds that, in making judgments, people will consider both stereotypic and non-stereotypic sources of information to arrive at their conclusion.

Although Bodenhausen and Wyer consider only the single complex case of judgments of guilt or innocence in a case of assault, they conclude that - at least in suchlike cases - the heuristic hypothesis is the most likely to apply. In this case, the effects of stereotype activation appear to override any other information available. The activation of a stereotype appears to be a simple matter: it can be performed by the simple suggestion of a name or ethnic background. Bodenhausen and Lichtenstein (1987) however presented evidence that effects may be to some degree situation-specific. Subjects presented with the complex task of judging the guilt or innocence of a defendant appeared to use ethnic stereotypes as an heuristic to simplify the task; whereas subjects given the simpler task of making a trait rating about the defendant took more account of the situation, thus supporting the integration hypothesis.

A conceptual replication of the Bodenhausen and Wyer study by Macrae and Shepherd (1989a) - using a different criminal situation and improved dependent measures - also tended to support the integration hypothesis. They conclude:

> "...the prediction that situational information would have little impact upon subjects' judgments when a stereotypical explanation for the crime was available was not supported... The present results...show that people are sensitive to situational information even when stereotypical explanation for the behaviour is available... That subjects did not ignore all the information available is perhaps not surprising given that they were instructed to pay close attention to all the information before making their judgments. The failure of previous studies to show the importance of situational information on

> causal judgements may be due in part to the experimental instructions presented to subjects." (Macrae and Shepherd, 1989a, p 324.)

Work in other, related fields also tends to support this hypothesis. Kelley (1971) presents a "discounting principle", by which it was proposed that perceivers make less extreme dispositional attributions where situational information is available; and this regardless of type of crime that the actor is said to have committed (and see also a wealth of research on causal attribution - Hewstone, 1990). Work by Locksley and her associates (Locksley, Borgida, Brekke and Hepburn, 1980; Locksley, Hepburn and Ortiz, 1982; Locksley, Ortiz and Hepburn, 1980) nominally demonstrates that the default hypothesis gives the only compelling account of the way in which social judgments are made. She finds that subjects' use of stereotypic information in making judgments is eliminated as soon as direct behavioural information about a target is provided. More careful consideration of her work, however, shows that her conclusions are based on inferences from just the sort of very simple judgment task that Bodenhausen, Lichtenstein and Wyer would predict to support the default hypothesis (see Grant and Holmes, 1981; Heneman, 1977). In each of Locksley's studies, the behavioural information provided has direct implications for the trait being judged (e.g., in the case of an honesty judgment, information about a theft might be given). Hence Locksley's evidence appears in fact to give more support to the integration

hypothesis than to the default hypothesis. Of course, the relationship between task complexity and reliance on stereotypic information may not be a simple one-to-one relationship, but may be mediated by such factors as the perceived consequences of an action (Harvey, Harris and Burns, 1975). Higgins and Rholes (1978) had subjects describe a stimulus person to another person who purportedly either liked or disliked that stimulus person. Subjects were found to shade their communication of the stimulus person's behaviour in the direction of the other's supposed beliefs. They conclude that abstract judgmental summaries are stored in memory along with a representation of stimulus information, and as this stimulus information decays over time, subjects increasingly rely upon this judgment to reconstruct the information. According to this hypothesis, therefore, the degree of integration which will take place depends upon the amount of time elapsed since encoding.

Work in support of the integration hypothesis can be tied in quite elegantly with a model proposed by Devine (1989a and b). Devine holds that, during socialization into their culture, everybody comes to learn the contents of that culture's stereotypes. Both bigots and non-bigots will unavoidably be aware of the contents of a particular culture's ethnic stereotypes. They will differ only in their personal beliefs about the actual applicability of these stereotypes: A bigot may consider the stereotypes valid, whereas a non-bigot may not. Where stereotypes

are activated, this will occur through an unconscious, uncontrollable process. The activation of personal beliefs however is likely to be a conscious, effortful process. The implication of this model is that, where information-processing demands on the perceiver are high, the cognitive effort is unlikely to be made to inhibit the activation of stereotypic beliefs. In this case then, (cf. Bodenhausen and Lichtenstein, 1987,) judgments are likely to rely more heavily on stereotypic information than on any other type of information. Where information-processing demands are low, however, stereotypic information is likely to have relatively little effect upon subsequent judgments.

This intuitively appealing model of the effect of stereotypes on judgments also supports the integration hypothesis: It appears likely that both stereotypic and non-stereotypic information *can* be used in forming a judgment; however, the relative importance of these two types of information will vary with the situation in which the perceiver finds him/herself. It should be borne in mind however that this model is applicable to an individual case only in a somewhat *post-hoc* manner: it is difficult to predict under what particular circumstances information-processing demands will be so high as to preclude the activation of personal beliefs in a judgment task. This lack of predictive power means that it will be difficult to produce very compelling evidence in favour of the theory.

Sherman, Judd and Park (1989) hold that there are multiple determinants of which items of stored knowledge will play a role in judgments. These are: the relative accessibility of different items of information; the similarity between features of stored items and incoming information; and the relative usefulness of different types of information in making judgments. The relative weightings of these different factors will usually vary with the situation; hence Sherman *et al*'s formulation can be conceptualised as a restatement of the integration hypothesis.

Judgments Made Outside the Juridic Situation

Srull and Wyer (1989) present the only paper in the literature to consider the effect of stereotypes on judgment outside the juridic situation. This is nominally a "model of impression-formation", but consists essentially of a list of fifteen postulates which attempts to account for a number of findings in the literature concerned with the effect of stereotypes on the encoding and organisation of, and memory for, social information. It is important to stress, however, that these postulates apply only to the situation in which a subject is explicitly attempting to form an impression about a target individual. Hence their applicability in the majority of real-world situations is questionable. However some of the postulates presented may nevertheless be applicable to the eyewitness situation.

According to Srull and Wyer, subjects asked to form an impression of a target person will first of all attempt to form an impression of him or her as generally likeable or unlikeable. This impression will be based on only a subset of the information received about that target: namely, the first information that is received. This, as Srull and Wyer note, implies a primacy effect on judgments.

Once such an evaluative impression has been formed, later behaviours will be interpreted in terms of that impression. Naturally enough, the ease with which this can be done will be a function of the evaluative consistency of the behaviour with the general evaluative impression.

If no evaluative impression can be extracted on the basis of the information initially received, subjects will tend to devote more attention to the behaviours they have already encoded to determine whether or not they have in fact interpreted this information correctly. This, presumably, will later give a recall advantage for this elaborately-encoded material (Bradshaw and Anderson, 1983). A corrolary to this is that, once an evaluative impression has been formed, inconsistent behaviours will receive more attention, in an effort to reconcile their occurrence. When asked to make a judgment about a specific characteristic of a target (e.g., his/her honesty) the subject will search memory for a representation of the target whose central concept relates directly to this characteristic (i.e., one which pertains specifically

to the honesty of the target). If a representation whose central concept has direct implications for a judgment cannot be found, the subject will retrieve and use the general evaluation-based person representation as a basis for this judgment (i.e., a target whose central evaluative representation is positive will probably be judged to be honest, and vice versa). It is important to note here that Srull and Wyer make no claims to the effect that any specific behaviour on the part of the target is retrieved from memory, just the evaluative impression. Hence this model makes no claim about the relationship between memory and judgment of the sort made by Bodenhausen (1988).

These last points especially may have implications for the eyewitness situation. An eyewitness is unlikely to have time to form any but the most basic evaluation of a criminal; and, naturally, this evaluation is likely to be negative. Hence, according to the Srull and Wyer model, when asked to make a judgment about, e.g., the appearance of the criminal in question, where the eyewitness has no specific memory of the feature of the criminal, recall will be based upon this negative, central evaluation. Hence such judgments are likely to be biassed in a negative direction, regardless of the actual appearance of the criminal. This implication of the Srull and Wyer model is quite intuitively appealing, reminiscent as it is of work on the "implicit personality theory" (Schneider, 1973; and see also Chapter 4).

Conclusions

Until recently, the literature on the effects of stereotypes on judgments was somewhat incoherent, lacking as it did any central focus or model to support or attack. Papers by Bodenhausen (1988) and Macrae (in press) have helped to rectify this situation, although it could still be argued that both concentrate on juridic matters almost to the exclusion of other matters, such as judgments of other people's dispositions, or physical characteristics, which may have more external relevance. An analagous charge could be levelled at Srull and Wyer (1988), since this paper is concerned essentially with impression-formation tasks.

Together with papers by Bodenhausen and Wyer (1985) and Bodenhausen and Lichtenstein (1987), the papers by Bodenhausen (1988) and Macrae (in press) hold that a perceiver's social stereotypes may interact in one of three ways with incoming social information to produce a judgment.

In the jury situation, the selective processing hypothesis (Bodenhausen, 1988) appears to give the best explanation of the effects produced in the laboratory. However, as Hamilton (1979) points out, the jury situation is likely to be a special case; involving as it does both "memory-based" and "on-line" components (Hastie and Park, 1986). Neither the biassed interpretation nor the heuristic hypothesis can be discounted. Research which addresses tasks

which are relatively more "on-line" and relatively less "memory-based" (i.e., research utilising paradigms which are more like real-world situations) must consider the effects of stereotypic information presented both at encoding and retrieval of stereotypic information. The eyewitness memory situation provides a good paradigm for such an investigation, as well as being an important area of research in and of itself. The following chapters consider the role of stereotypic information in the creation of judgments relevant to this area: violent/aggressive actions; facial and whole-body information.

Three hypotheses have been proposed to account for *when* stereotypes will affect judgments made. The preponderence of evidence appears to be in favour of the integration hypothesis, which holds that both stereotypic and non-stereotypic information is likely to be used when a judgment is made. Probably, the weight which is given to each type of information will vary with circumstance, and particularly the information load with which the perceiver is asked to deal (Locksley *et al* 1980a, b; 1982; Devine 1989a, b). The "average" eyewitnessing situation appears intuitively likely to present the perceiver with a great deal of information all at once; therefore one would predict a powerful effect of stereotypes on judgment in this situation. This accords with the predictions of Srull and Wyer (1989) for the

impression-formation paradigm. It is with this area that the following three chapters aim to deal.

CHAPTER 3 - DOES SIZE MATTER? THE EFFECT OF BODY-BUILD ON THE JUDGMENT OF AN AMBIGUOUS ACTION

"The bigger they are the harder they hit."
Woody Allen

"...not only huge, but horrible huge."
Ken Kesey: Little Tricker the Squirrel Meets Big Double the Bear

Introduction

The aim of the present experiment is twofold. First, by considering the effects of stereotyped characteristics on the interpretation of behaviour, it acts as a starting point for the investigation of the effects of stereotypes in the eyewitness situation. Second, it attempts to further present knowledge of body-build stereotypes. To address the first issue: Some, (e.g., Reiser, 1976) still hold with the so-called "video-recorder" theory of memory which holds that all incoming information is stored in its original form in memory throughout life. This is however hardly the generally accepted view. The malleability of human memory has been a subject for psychological research at least since Bartlett (1932) conducted his first experiments into the phenomenon. Among the factors which may be considered particularly important in shaping memory are social factors: e.g., generic schemata for social events (see Chapter 6), or social stereotypes. The powerful implications that stereotypic information can have for social memory have been demonstrated in a number of studies. Principle among these are those presented by Hastie (1981); Hastie, Park and Weber (1984); Rothbart, Evans

and Fulero (1979); Snyder and Uranowitz (1978c); Srull (1981); Stangor and Ruble (1989); Wyer, Bodenhausen and Srull (1984); and Wyer and Gordon (1982/84). All of these studies are reviewed in Chapter 6. Stereotypic information is likely to impact upon the eyewitness situation in at least two fundamentally dissimilar ways (discounting for the moment the possibility of a third sort of impact - that of the retrieval schema - also discussed in Chapter 6). The nature of a witnessed crime may bias memory or interpretation of a stereotyped aspect of the person or persons involved (e.g., facial features); alternatively, perception of a particular stereotyped aspect of a target (e.g., some aspect of the facial features) may bias memory or interpretation of the crime itself. These are two sides of the same coin: each could, and probably does, have implications for real-world situations. The former circumstance comes into effect where, for example, a person witnesses a particularly violent crime in which witnessing conditions are sub-optimal and the witness is unsure about, say, facial or body features of the criminal in question. At an identification parade, or while constructing a Photofit, the witness' memory of the violence of the crime might make him/her misremember the criminal as having a more stereotypically "violent" look than s/he otherwise would have done (a broken nose or cauliflower ears, perhaps): leading to the arrest or imprisonment of the wrong person. The latter circumstance comes into effect where, for

example, viewing conditions are somewhat better, and the witness is able to perceive the criminal as in fact being possessed of a stereotypically "violent" look (really having a broken nose or cauliflower ears). In this instance, the witness might misremember the incident as more violent than s/he would otherwise have done: leading perhaps to a longer jail sentence than would otherwise have been given. Evidently, it is necessary to consider each of these possibilities in an investigation of the effects of stereotypes in eyewitness situations. It is the latter effect which will be investigated in the present experiment.

Such an experiment could conceivably utilise either body or face stereotypes. Whole-body stereotypes were however chosen. There were two reasons for this: 1) body stimuli may be simpler in nature and easier to manipulate than face stimuli, since they appear likely to vary along a smaller number of continua; 2) relatively little research has addressed the issue of body-memory or judgment; hence to conduct an experiment using whole-body stimuli has greater marginal utility than to conduct another experiment using face stimuli; it provides an opportunity to increase scientific knowledge of these potentially important stimuli.

Literature Review

The present study is less concerned with the objective relationship between physique and psychology, as the

perceived relationship. It concerns the way in which a stereotypic body-build might affect the judgment of an action. To this end, it is necessary to review here the state of the literature on whole-body stereotypes.

Whole-Body Stereotypes: To the extent that everyone is a "lay psychologist", everyday perceptions of character are likely to be influenced by what is seen - in this case, whole-body information. To what extent then are social judgments mediated by whole-body stereotypes?

Much of the work on this question has its basis in the research of Sheldon (1927). Sheldon developed the idea that the human physique was determined by three independent variables, which he termed endomorphy, mesomorphy and ectomorphy (although this idea was not strictly original to him, being similar to work that had been carried out by psychologists and medics from Hippocrates to Kretschmer - see Sheldon *et al*, 1940 pp 11 - 15, and Eysenck, 1947). Endomorphy referred to the relative predominance of "soft roundness" in the body; this occurred when the digestive viscera had greatest dominance of the bodily economy. Mesomorphy referred to the relative predominance of muscle, bone and connective tissue, giving the body a generally more rectangular appearance. Ectomorphy referred to the relative predominance of "linearity and fragility" (Sheldon *et al*, 1940 pp 5 - 6).

Sheldon's system was a means of classifying any human physique on an index of three digits, ranging in value from 1 to 7. The first digit referred to the relative contribution of endomorphy to that body; the second to the relative contribution of mesomorphy; the third to the relative contribution of ectomorphy. In each case, the digit 1 referred to the smallest possible contribution, the the digit 7 to the largest. For instance, a 7 1 1 physique would be characterized by extreme endomorphy and minimal mesomorphy and ectomorphy. A 1 1 7 would be characterized by extreme ectomorphy and minimal endomorphy and mesomorphy. A 4 4 4 would lie at the mid-point of all these scales.

Sheldon's classification system is unfortunately of limited use to the researcher into eyewitness memory. It takes little account of the factors which might be thought most salient when viewing another person: factors such as height and weight, and the relative length of the limbs. And while a witness to a crime might be able to make a fairly accurate judgment of the perpetrator's relative endomorphy or mesomorphy, say, there is little reason to believe that s/he would be able to place the perpetrator's physique with any degree of accuracy on Sheldon's scales. Additionally, Sheldon's work has more recently come under fire for being insensitive to subtle variations in physique (see, e.g., Powell *et al*, 1974). It fails to provide any really useful system for the generation of body-shapes in the way that Photo-fit, say, provides a system for the generation of facial stimuli.

Nevertheless, Sheldon's system has provided the basis for a number of studies in the psychological literature. Many of these imply that different social stereotypes may exist for different body-builds. Some of the most important are mentioned here.

Lerner and Korn (1972) studied the development of body-build stereotypes in males from different age-groups between five and twenty years. They found that, at all ages, the mesomorph was the most positively regarded somatotype; and that there was a generally negative view of the endomorph and a slightly less negative view of the ectomorph: in other words, value judgments varied with physique presented. Yates and Taylor (1978) had thirty subjects attribute sixty personality traits to outline drawings based on Sheldon's three primary somatotypes. Somatotypes were not only found to be strongly stereotyped, but stereotyped in accordance with Sheldon's predictions: a rare finding in support of his theory. Hiller (1982) had college students write stories concerning either normal or overweight targets, and found that more overweight targets were associated with more negative personality characteristics. Iwawaki and Lerner (1974) had subjects of each sex attribute thirty behavioural descriptions to pictures of male endomorphs, ectomorphs, and mesomorphs. Each sex was found to give mostly positive evaluations of mesomorphs, and mostly negative evaluations of others. Strongman and Hart (1968) found that subjects considered the mesomorphic somatotype to be associated

with "competitive aggressiveness" and "assertiveness of posture and movement", while the endomorphic somatotype was associated with neither of these characteristics, and the ectomorph was seen as an introvert. Gacsaly and Borges (1979) had subjects attribute twenty-four personality traits to one of six body-types, varying by height and somatotype. The tall mesomorph was attributed the most socially desirable personality traits, and the endomorphs the least socially desirable personality traits. A positive image was found for ectomorphs in this study. This finding was mirrored by Ryckman, Robbins, Kaezor and Gold (1989), who had subjects attribute a variety of traits to male and female target mesomorphs, ectomorphs, and endomorphs. They found a positive stereotype of mesomorphs, and a negative stereotype of endomorphs. They also found ectomorphs to be viewed in a generally favourable light; an unusual finding in the area, and one which might conceivably reflect changing societal views about the kind of body-shape that is desirable: a phenomenon which may also be seen in the growing emphasis on slimming and increasing prominence of the "super-model".

Rappoport (1975), in a slight break from traditional research in this area, considered male body-height stereotyping. She found that stereotyping existed for all subjects, although field-dependent subjects attributed more positive traits to a tall target than did field-independent subjects. This was true for

both tall and short subjects. Women on the whole were found to stereotype less than men.

Biases in Body-Perception: On the basis of the studies mentioned above, it seems likely that everyday social perceptions are mediated to some extent by stereotypic judgments made on the basis of physique. The picture is further complicated by the systematic egocentric biases which also appear to be involved in body-perception. Estimates of others' heights and weights are notoriously inaccurate. In the US criminal case People vs Thomas, a witness had described two perpetrators as being each about 5 foot 8 inches tall, and then went on to pick defendants of 6 foot 5 inches and 6 foot 4 inches from identification parades (Parker, 1973, p 256). Errors are made even by trained observers (Clifford and Richards, 1977).

Subjects tend to use their own height and weight when estimating the height and weight of others (see, e.g., Bailey, Shinedling and Payne, 1970; Dunaway, 1973; Gorchinski, 1973; Williams, 1975). Flin and Shepherd (1986) extended the findings of earlier studies by using a range of targets rather than just one individual. They had targets ask members of the public for directions in a busy city centre. Once the target had disappeared from view, subjects were asked to estimate his height and weight. Both estimates varied systematically according to the height and weight of the target and, to a lesser extent,

subjects' own height and weight. Importantly, Flin and Shepherd also concluded from their study that eyewitness errors were related to initial misperceptions of the target's physical characteristics, rather than to any subsequent memory distortions caused by the body-build of the interviewer. Unfortunately, it is unclear which is more important: the subject's *actual* or *perceived* height and weight.

This is important because of the general inaccuracy of people's perceptions of their own height and weight. This effect is especially pronounced for overweight people and those who suffer from eating disorders, but is by no means restricted to them (see, e.g., Glucksman and Hirsch, 1969; Bailey, Shinedling and Payne, 1970; Gardner, Martinez and Sandoval, 1987; Collina, McCabe, Jupp and Sutton, 1983 - who found obese subjects to overestimate their own body-size by a mean of 19%; and Collins, 1987, who found obese subjects to overestimate their own body-size by amean of 12%, as against a mean of just 0.88% for normal-weight control subjects). Schonbruch and Schell (1967, cited in Collins, 1987) summarised this literature in concluding that persons with deviant physiques made more errors in judging body shape than persons with more normal physiques.

Not that these effects are limited to those with "deviant" body-builds. Systematic distortions are to be found throughout the population. Singer and Lamb (1966 - cited in Collins and Plahn, 1988, p 320) found

that most female adolescents systematically distorted self-estimated physique toward their estimate of their ideal physique; and Collins and Plahn (1988) found that females tended to underestimate their own body dimensions, while males tended to overestimate them. Minahan (1971) found that teenage girls tended to perceive their figures as more attractive than they in fact were. Given that anorexia is most likely to develop during the teenage years, it is obvious that prediction of the direction of any distortion of self-body-image during this period will be extremely difficult.

Despite studies such as those reviewed above, the issue of whole-body perception and stereotyping has consistently failed to receive the attention that its social psychological importance warrants. A pair of papers by Powell, Stewart and their colleagues highlight the basic work that still remains to be done.

Powell, Stewart and Colleagues: Powell, Tutton and Stewart (1974) investigated how different two physiques needed to be in order to be stereotyped differently. They had fifty subjects rank six physiques (two examples of each of Sheldon's somatotypes) from "most suiting" to "least suiting" each of fifteen concepts. It was discovered that "superficially similar physiques, members of the same overall somatotype, are clearly differentiated in terms of stereotype" (Powell *et al*, 1974, p 422).

A follow-up paper by Litman, Powell and Stewart (1983) made a more controversial claim. They argued that the original results may have been an artifact of the ranking procedure employed in the experiment. Therefore they ran a conceptual replication of the experiment, this time not using the ranking procedure, but simply having subjects rate targets, one at a time, against their own internal standards. Subjects were presented with a target, plus six seven-point bipolar scales on which to rate it. Again, clear evidence was found for systematic differential stereotyping of similar physiques. Twenty-three of the twenty-four differences found favoured the less extreme target.

Litman *et al* claim that the categorical view of whole-body stereotyping (as adopted in the papers discussed above) is jeapordised by these findings. It would, they claim, require a huge number of stereotype categories to account for all possible whole-body types. Furthermore, person-perception and stereotyping may not necessarily be independent processes. They may be two sides of the same coin; the particular process employed in a given experiment depending on the stimuli employed. Where stimuli are emotive and logically indefensible, the process will be called "stereotyping"; where they are realistic and reasonable, it will be called "person perception".

The potential ramifications of this paper for whole-body stereotyping research are evident. It is possible that while many researchers have believed

themselves to be investigating the widespread influence of whole-body stereotypes, they have really been investigating the implications of their own target stimuli: hence their conclusions may have less external validity than has been claimed for them. Within the present field the ramifications are no smaller: the eyewitness situation is after all a real world situation, involving complex, subtle, and meaningful stimuli. Rigorous experimentation on eyewitness memory for whole-body information would provide a possible means of testing the claims of Powell, Stewart, and colleagues.

However, it is possible to argue that the results achieved in these papers do not justify the conclusions reached. From the finding that whole-body stereotypes do not map onto somatotypes in a one-to-one fashion, Powell, Stewart and colleagues argue that such stereotypes may not exist. Certainly, their results may justify the claim that whole-body stereotypes are more fine-grained than earlier studies would lead one to believe; but it is a big step from there to the claim that there is no qualitative difference between body-stereotyping and perception. It is possible to claim that the attribution of stereotyped characteristics correlates with body-build without claiming a direct one-to-one relationship between classes of somatotype and stereotype categories. Indeed, such a model would fit the results both of Powell, Stewart and their colleagues, and of Lerner and Korn, Yates and Taylor, etc.,

discussed above. If one were to hypothesize, say, the trait of introversion to be stereotypically linked with the ectomorphic somatotype in such a way that the more ectomorphic a target appears, the more introverted he is likely to be judged, this would account for both types of result. Where a study uses just one example of each body-type, evidently the ectomorph is likely to be judged more introverted than the endomorph or the mesomorph. Where two or more examples from each category are given, then more ectomorphic targets will be judged more introverted; a result analagous to that achieved by Litman, Powell and Stewart. This seems an intuitively appealing conclusion regarding the everyday use of whole-body stereotypes.

Pilot Study

INTRODUCTION The aim of this study was to investigate the relationship between perceived size of a male target's body, and his perceived aggressiveness. Everyday experience, not only of real-world interactions but also of, e.g., Hollywood movies, implied a stereotypical positive relationship between these variables. It was hypothesized that a significant positive correlation would be established between perceived body-size and perceived aggressiveness.

MATERIALS Ninety-six still photographs were taken of male targets using a Pentax Asahi SP1000 camera loaded with Kodachrome Gold 400 film. All targets were

photographed against a plain white wall, with heads and facial features occluded.

SUBJECTS Ninety-six volunteer subjects were used in this experiment. Forty-six subjects were male and fifty were female. All subjects were aged between 18 and 27 years. Each experimental session took a maximum of five minutes. None of the subjects was paid for participation.

PROCEDURE Each subject rated five photographs on nine-point scales on a number of variables: honesty, intelligence, attractiveness, friendliness, co-operativeness, and aggressiveness. Correlations were calculated between these character traits and the perceived body-size of the targets in the photographs. Each photograph therefore received five ratings on each scale. The exact definition of "body-size" was left to each individual subject to determine; if they enquired, however, they were told both height and weight should be considered.

RESULT A correlation of $r = +0.309$ was established between rated aggressiveness and body-size across the targets, $p < 0.001$ on a one-tailed test. No other correlation was significant.

CONCLUSION This provides support for the theory that judgments about aggressiveness correlate with judgments about body-size. The result also indirectly supports the theory proposed above, that whole-body stereotyping is unlikely to be a categorical process, but that the attribution of stereotyped characteristics correlates with body-build. All other

things being equal, targets of larger perceived body-size are likely to be stereotyped as more aggressive than those of smaller perceived body-size. This implies that whole-body stimuli might be utilised in a study in which the variable of aggressiveness was studied. This variable has particular relevance to the crime situation. The body-size variable might be used in the same way that the race variable was used in an experiment by Duncan (1976), discussed below.

Heider and Simmel (1944)

This important paper provides an early investigation into factors which can affect subjects' judgments of ambiguous events. In this case, the factor in question was the size of geometric shapes. Heider and Simmel had subjects watch a complex sequence of actions involving numbers of geometrical shapes, including a large triangle (T), and a small triangle (t). Subjects' interpretations of the behaviours being performed by the geometrical shapes varied partially with the size of the shape - e.g., when t moved behind T, t was said to be "following" the larger triangle; but when T moved behind t, it was said to be "chasing" the smaller triangle. Subjects also attributed human characteristics to the shapes in trying to explain their behaviour: e.g., T was troublesome; t was heroic. If such an effect could

hold for artificial stimuli, it seemed reasonable to suppose that it might also hold for human actors.

Duncan (1976)

In a sense, this experiment constituted an application of the conclusions reached by Heider and Simmel to a more realistic setting: i.e., one in which human targets were involved. Obviously, the aggressiveness of a target person is an important factor in the present field of study. A paradigm allowing investigation of the effect of body-size (and the correlated stereotype of aggressiveness) on the interpretation of an ambiguous incident would therefore have great interest.

Duncan presented subjects with brief video-recordings ("video-clips") of target persons discussing "risky-shift" dilemmas (Wallach, Kogan and Bem, 1962). The discussion featured in each clip became more and more heated, until one discussant gave the other what Duncan describes as an "ambiguous shove", i.e., one that was neither obviously violent nor obviously playful. The race of target persons was varied across conditions: both "shover" (or "protagonist") and "shovee" (or "victim") could be either black or white: giving a total of four basic experimental conditions. Each video-clip was interrupted four times. Subjects were given a ratings form based on the Interaction Process Analysis form devised by Bales (1979), and asked to use this to rate the behaviour of the target who was acting immediately

prior to each interruption. The rating of interest was the final one, i.e., the one immediately following the "ambiguous shove".

Duncan found that the shove was more likely to be perceived as "violent" when perpetrated by a black actor than when perpetrated by a white actor, likely to be rated as "more intense", and also that subjects were more likely to attribute the cause of the behaviour to situational variables in the case of a white protagonist and to person variables in the case of a black protagonist. Strangely, although he gave subjects the opportunity to rate the behaviour as "aggressive", Duncan failed to test the effect of race on the incidence of use of this category.

Whites, therefore, were thought to have committed the act because of the situation in which they found themselves; whereas blacks were thought to have committed the act because of the type of people they were. Duncan concluded that "the threshold for labelling an act as violent is lower when viewing a black committing the same act" (sic, p 596).

This finding is explained in terms of the stereotype of "blacks as violent" acting such as to lower the perceptual (judgmental) threshold for violence in black actors.

Duncan devotes some space to discussion of the attribution data. He had hypothesized that subjects would attribute violent behaviour - in all conditions - to stable personality traits of the actor. However, as noted above, this occurred only where the

protagonist was black. Duncan explains this by reference to the suggestion made by Jones and Nisbett (1971) that "the illusion that our *reactions* are *perceptions* is sustained in part by the apparent consensus accompanying most of our reactions, a consensus that may rest as much on transmitted cultural norms as on the compelling features of objective 'reality.'" (Duncan, 1976, p 597, my italics). Hence, not only dispositional, but also situational attributions are strongly affected by the action of stereotypes. The term "situational attribution" (as used here at least) may indeed be a misnomer; to judge by Duncan's conclusion, "situational attributions" seem to be a special type of "dispositional attribution" (subjects appear to have applied to certain actors a stereotype of the sort "the type of person likely to be influenced by situational factors"). This will come as no surprise to students of the fundamental attribution error.

Duncan's conclusion may be worth repeating here, since it ties in his study with the subject matter of the present thesis:

> "It would appear that the black man is imbued (stereotyped, categorized, etc.) with such salient personality properties (e.g. given to violence) that these traits tend to engulf the field rather than be confined to their proper position, the interpretation of which required additional data about the situation. Dispositions then are treated as causal and are "packaged". Cronbach (1955) and Mischel (1968) recently spoke of these packages as an implicit personality *theory*, an intuition about how traits interact."
> (Duncan, 1976, his italics).

Sagar and Schofield (1980): Sagar and Schofield held that Duncan's study raised more questions than it answered. They wanted to discover whether the phenomenon would generalise to black subjects: Duncan having used only white subjects in his study.

In an effort to generalise from Duncan's materials to other materials, they gave their subjects (black or white sixth-grade children) oral descriptions and artist's renditions of four different dyadic interactions: bumping in the hallway, requesting food from another student, poking a student in the classroom, and using another's pencil without asking. Four different sets of stimuli were used, each depicting one of the four possible pairings of black and white protagonist and victim. Subjects rated the behaviour occurring in each picture on four traits (playful, friendly, mean, threatening), using seven-point scales in each case. They then rated the personality characteristics of protagonist and victim on sets of seven-point scales (e.g., thoughtless-considerate, strong-weak).

Sagar and Schofield achieved parallel results to Duncan: i.e., behaviours performed by black actors were perceived as more "negative" (i.e., more mean/threatening) than identical behaviours performed by white actors; although the incidence of "positive" (i.e., playful/friendly) ratings did not vary with condition. This effect obtained for both black and white subjects. Ratings of personality were heavily

influenced by the depicted behaviour and the stimulus person's role: protagonist or victim. Protagonists were judged to be ruder, meaner, more thoughtless, threatening, unfriendly, and less likeable, than victims.

Sagar and Schofield conclude: "...in the existing social order, the stereotype is all too real. To activate it, the person engaging in an ambiguous behaviour need only be black" (Sagar and Schofield, 1980 p 597).

A Paradigm for Investigation of the Effect of Body-Size Stereotypes on Judgment of an Action: Given the relevance of Duncan's design to the present situation, and validation of his results by Sagar and Schofield, it was decided that a similar paradigm would be appropriate for investigation of the effects of perceived body-size on the interpretation of actions. Particularly appealing was the fact that the paradigm had already demonstrated an effect of social stereotypes on judgments. Hence it might be expected to demonstrate the effect again if indeed it was present: perceived body-size should affect the judgments made about ambiguous actions.

Experiment 1

The hypotheses of the present experiment were:

1) large perceived body-size should cause similar reactions to black skin-colour in Duncan's experiment. In other words, an "ambiguous shove"

administered by an actor with a large perceived body-size should be viewed more negatively than a similar shove administered by an actor with small perceived body-size. It could be perceived as more violent (as in Duncan's experiment), more aggressive, or, combining the ratings on these two scales, more violent/aggressive. The actions of a perceived larger target should also be judged as more intense than those of a smaller perceived target, although this is not central to the experiment: it is not a necessary condition of a stereotypic action that it is perceived as more intense than a non-stereotypic action. The stereotypic action of the larger perceived target would be more likely attributed to stable personality traits than to external forces or the situation in which the actor found himself.

2) smaller perceived body-size should cause similar reactions to white skin-colour in Duncan's experiment: i.e., reversing the stereotype of targets with a larger perceived body-size.

The effects of the actual body-size of the actors was also considered, although this manipulation was not central to the experiment. Subjects' perceptions of the actors' body-sizes was the more important measure; this manipulation acted to check whether the experimental manipulation of body-size was the same as subjects' perceptions of this variable.

Methodology

Design: To ensure maximum validity, the design of the present experiment closely paralleled that of Duncan (1976), with the exception that body-size rather than race of target was varied. Otherwise, changes in experimental paradigm were limited. In Duncan's experiment, subjects were misled into believing that they were actually witnessing, via closed-circuit television, an ongoing interaction between two other experimental subjects. In the present study, however, subjects were led to believe that they were witnessing a video-recording of an interaction which had taken place between two experimental subjects earlier in the week. This change was made for two reasons: partly because Duncan's original paper did not specify what function the manipulation was supposed to serve, and partly because of the simple fact that the layout of the St Andrews School of Psychology did not permit copying of the original set-up. The one conceivable advantage of Duncan's approach is that subjects believed themselves to be watching a real interaction, rather than a pair of actors plying their trade. However, it is believed that this problem was circumvented by telling subjects at the time of the experiment that they would be viewing an encounter video-recorded previously. This approach had the advantage that subjects did not find themselves in the melodramatic position of apparently viewing a "live" interaction dissolving into an aggressive

confrontation which had to be broken up by the experimenter as in Duncan (1976, p 593).

Another break from Duncan's procedure was to provide subjects with photographs and brief (fictional) biographical details about targets in the experiment. The purpose here was to impress upon subjects the divergence or similarity (as the case happened to be) of the body-sizes of the targets. This was a necessary break with Duncan's procedure: whereas race is an instantly recognisable characteristic even when a target is seated at a table, body-size is somewhat less so.

Materials: The stimulus material was a videotape of two of a total of four actors (chosen to be divisible into two pairs on the basis of height and weight) discussing a situation derived from the "dilemma of choice" or "risky shift" item list constructed by Wallach, Kogan and Bem (1962). The actors were four white males, all aged between twenty-five and twenty-seven years. Two actors, used as "large" targets, were each approximately six foot one inch tall, and weighed approximately one hundred and seventy-five pounds. The others, used as "small" targets, were each approximately five foot ten inches tall, and weighed approximately one hundred and fifty pounds. This relatively small absolute difference in body-size was employed since this was believed to make the experimental materials more realistic. No actor had particularly distinctive or obtrusive features. All

were casually dressed, as in everyday life. The dilemmas used were the same as those used by Duncan (1976): i.e., numbers 1 and 2 from the list provided by Wallach *et al*. The discussion of interest was that which centred around dilemma number 1 (see Appendix 5 for an account of this dilemma, and also a copy of the script employed). The warm-up item (i.e., the first video-clip watched by subjects,) was number 2 from the list, concerning a man with a heart ailment who must choose between changing his way of life or undergoing a potentially lethal operation.

These video presentations were constructed in two stages: 1) a few days before the recordings were to be made, three actors came to the School of Psychology to improvise discussion of the dilemma of choice situations. A number of these improvisations was conducted, and all were tape-recorded. Tapes were later played back to determine which led most convincingly to the altercation culminating in the "ambiguous shove". This tape was then transcribed in the form of a script. 2) The four actors involved learned the script, and performed it on videotape.

Following Duncan's own procedure, the comparability of behaviours depicted in the video-clips was tested by having independent subjects rate the personalities of the actors involved in each clip. Both actors in each clip were rated by eight independent subjects on thirteen character traits, using seven-point scales. These data were subjected to analysis of variance. For no character trait was an effect of video-clip

found: indicating that actors did not act in significantly different manners across clips (Appendix 1, Tables 1 - 12). Hence inter-clip similarity was comparable to that in Duncan's experiment.

The ratings forms used in this experiment contained a simplified version of the Interaction Process Analysis (IPA, Bales, 1970). The ratings system was simplified in the same way as in Duncan (1976) - i.e., the number of ratings categories was reduced, and only periodic, rather than continuous, ratings had to be made.

Instructions for use were presented with each IPA ratings sheet. These read:

> "The observer should make one rating each time there is a pause in the video-recording. Upon receiving the signal the observer should rate the behaviour which took place immediately prior to the signal. In making the rating the observer should do, in order, the following things: (1) decide which person (A or B) was emitting the behaviour to be rated and then circle either A or B at the top of the rating sheet; (2) decide into which of the 10 major categories the behaviour falls and circle the number of that major category; (3) indicate whether the behaviour was of relatively low intensity or high intensity by making a check mark on the scale to the right of the major category being used; (4) fill out the three subscales under the major category being used. As soon as all these steps have been completed, turn over the filled-out rating sheet and get ready for the next rating signal. If the behaviour to be rated does not fall into one of the ten categories provided, circle # 11 (Uncategorizable) and briefly describe the behaviour. Use this category, however, only as a very last resort. It is important to note that each rating should be done on a separate sheet and only one person (A or B) and only one category of behaviour (one of

> the ten categories) are scored on each rating. Also, a rating is to be made only when the signal is given."

The ten major categories used (as in Duncan, 1976) were: dramatizes, gives information, gives opinion, gives suggestion, asks for information, asks for opinion, asks for suggestion, playing around, aggressive behaviour, and violent behaviour. Each category name was accompanied by a nine-point intensity scale, and three minor categories ("subscales"), each presented with its own nine-point scale. Aggressive and violent behaviour categories had the subscales provoked - unprovoked, intentional - unintentional, and consequential - trivial. Dramatizes and playing around categories had subscales emotional - calm, enthusiastic - unenthusiastic, and show-off - modest. Gives information, gives opinion, and gives suggestion were accompanied by subscales tactful - tactless, intelligent - unintelligent, and autocratic - democratic. Asks for information, asks for opinion, and asks for suggestion were accompanied by subscales inhibited - uninhibited, demanding - undemanding, and soft - loud. The rated intensity of the behaviour of the targets was calculated by combining the ratings administered by subjects on the major scale chosen and its associated sub-scales.

Subjects: Subjects were forty white undergraduates and postgraduates from the University of St Andrews, aged between eighteen and twenty-seven years. Twenty-

one subjects were female, nineteen male. All were recruited through a sign in the School of Psychology offering £2 for participation in a study of "interpersonal behaviour research". Only one subject participated in the study at a time. Each subject was randomly assigned to one of the four experimental conditions.

Procedure: On arrival, each subject was escorted to the Social Psychology Laboratory. Here s/he found a desk and chair at a comfortable distance (approximately two metres) from a television and video-recorder. On the desk was a copy of the IPA ratings system (see above).

The subject was told that, earlier in the week, independent subjects had been invited to the School of Psychology to discuss a number of simple issues in front of the camera.

The subject was then given instructions very close to those given by Duncan to his subjects. The subject was told:

> "The instructions for the use of these materials are given at the top of the sheet. In a moment I will give you time to read them and ask any questions you may have about them. First I want to tell you a little about this ratings system. This study is part of a project aimed at developing a new system for rating of interpersonal behaviour. As you can see, this new system involves quite simple categories of behaviour. This is because it is intended for non-professional observers. In the recordings of the interactions which you are about to watch, there will be a number of points at which the screen will

> go blank. At these points I want you to complete your ratings. You won't have a lot of time, so respond quickly. Now you will do ratings of two interactions. The first is by way of practice. I will stop the video-recorder after the first practice series to answer any questions you might have. Please do not ask any questions until the end of the first interaction."

Design: Each subject was randomly assigned to one of the following experimental conditions (clips): large protagonist - small victim; small protagonist - large victim; large protagonist - large victim; small protagonist - small victim.

Each videotape contained two interactions. The first "warm-up" risky shift situation acted as the subject's "practice rating" - see above. As in Duncan's study, the subject was given two practice ratings during this interaction (which lasted approximately five minutes). After the practice run, the experimenter checked the subjects' responses to ensure that the rating system was understood and was being used correctly. If the experimenter was satisfied that this was in fact the case, he restarted the video-recorder to show the subject the second (experimental) video-clip.

This interaction was also approximately five minutes long. Immediately after the "ambiguous shove", the screen went blank, signalling the subject to rate the interaction. This was the subject's second rating during this particular clip.

The experimenter collected the subject's completed ratings forms and asked him/her to complete a short questionnaire. This questionnaire was introduced in the following manner:

> "Pick out a sequence of behaviour which left an impression on you and answer the questions on the form as they pertain to that interaction. Better yet, why don't you use the last sequence of behaviour which you rated [reminding subjects of the altercation in which A shoved B]."

This wording was used, as in Duncan (1976), so as to give the impression that any behaviour sequence would do, but the experimenter had made a spur-of-the-moment decision to use the final interaction. A number of nine-point scales was presented in the questionnaire, concerning a) extent to which observed behaviour should be attributed to external forces, b) extent to which it should be attributed to the actor as a person, c) extent to which it should be attributed to the issue discussed (stimulus) and d) other (this category was left open for interpretation by the subjects, as in Duncan's experiment, but could be interpreted as meaning some combination of the a, b, and c, as suggested by Duncan). Finally, subjects were asked which person in the video-clip they believed to have been physically larger. If they could remember which target (if either) was said to have been larger, they were asked to give this answer; otherwise they were asked to guess on the basis of their perception of the video-clip.

After completing this questionnaire, subjects were fully debriefed by the experimenter and paid.

Results

The results of this experiment fall into two discrete categories. Results for perceived and for actual body-size are presented separately.

Perceived Body Size: The raw data by perceived body-size are given in Table 3.1. It can be seen from this table that there are unequal numbers of subjects per cell, and also that there are only three data-columns - for equal, large-small, and small-large pairings. This is because different numbers of subjects perceived the body-size pairings in each of the three different ways. Although two sets of equal body-size pairings were used (two large actors in one pair, two small actors in the other) subjects obviously could not be aware that a pair of targets that appeared of equal size were in fact "large" or "small" targets. Only four behaviour categories - playing around, dramatizes, aggressive behaviour, and violent behaviour - were used in the rating session. This pattern of responses is identical with that found by Duncan (1976). 82.5% of the ratings (thirty-three of forty) used the categories "aggressive" or "violent". Twenty-three ratings (or 57.5% of the total) fell into the "aggressive" category.

Major Category	Equal	Large-small	Small-large
Playing Around	2	0	0
Dramatizes	4	0	3
Aggressive Behaviour	2	14	7
Violent Behaviour	0	6	2

TABLE 3.1 - MAJOR CATEGORY FREQUENCIES BY PERCEIVED PERPETRATOR/ VICTIM BODY-SIZE PAIRINGS

Chi-square revealed that perceived body-size pairings significantly affected the number of "violent" judgments made ($\chi^2(1) = 4.0688$ $p < 0.05$). A greater number of "violent" judgments was made in the perceived large protagonist/small victim condition than the perceived small protagonist/large victim condition, or the perceived equal size condition. This confirms one of the most important hypotheses of the present experiment. Although body-size pairings were not found to affect statistically the number of "aggressive" judgments made ($\chi^2(1) = 2.8232$ $p > 0.05$), a similar trend can be seen in the raw data. Perceived body-size was also found to affect the number of playing around/dramatizes judgments made (χ^2

(2) = 14.8036 $p < 0.01$), although one must be cautious in interpreting this, because of the small numbers involved.

When aggressive and violent ratings were combined, chi-square remained significant: $\chi^2(2) = 6.8920$ $p < 0.05$. Except in the single case of "aggressive" judgments (which shows a similar trend but is not statistically significant), all of these results conform to the experimental hypotheses, above; with the implication that perceived body-size may affect judgment of an action in a manner parallel to that of race in Duncan's original experiment.

When results were broken down by body-size pairing, two pairings were found to affect judgments made. Where perceived body-size was equal, $\chi^2(1) = 6.175$ $p < 0.05$, and where a large protagonist was matched with a small victim, X^2 (1) = 5.7318 $p < 0.05$. In the case of equal perceived body-sizes, six of eight ratings fell at the playing around/dramatizes (or more "positive") end of the spectrum; whereas in the case of a large perpetrator and small victim, all twenty ratings fell at the aggressive/violent (or more "negative") end. With a small perpetrator and a large victim, $\chi^2(1) = 2.4535$ $p > 0.05$: the distribution of ratings judgments in this condition appeared more evenly spread. These findings are all in accordance with the hypotheses: a larger perceived protagonist was expected to be judged more negatively than a smaller perceived protagonist or one whose body-size was identical with that of his victim. Where body-

sizes were perceived to be equal, ratings were expected to fall at the more positive end of the spectrum. No specific hypothesis was advanced for the case of a perceived small protagonist and large victim.

The next thing to be considered was perceived intensity of the action. The raw data generated here are presented in histogram form in Figure 3.1. Perceived intensity of actions was calculated by combining the intensity ratings of the major categories and subscales used in rating the behaviour of the targets at the point where the ambiguous shove was administered. These ratings were then subjected to analysis of variance. No effect of perceived body-size pairings was found: $F(2,37) = 1.2660$ $p = 0.2939$, and there was no interaction between perceived body-size pairing and intensity of particular behaviour rated: $F(6,111) = 1.1187$, $p = 0.3561$ (see Appendix 2, Table 1). This ran counter to hypothesis, although, as mentioned above, this hypothesis is not central to the experiment.

Attributional behaviour: In the questionnaire administered at the end of each session, subjects were asked four attributional questions concerning the behaviour of the protagonist on the video-clip (see above). It was predicted that when the protagonist had a larger perceived body, his aggressive behaviour would more likely be attributed to stable internal forces, and less to varying external forces, than when

he had a smaller perceived body. Attributions made for the various perceived body-size pairings were examined using analysis of variance. The raw data are presented in graph form in Figure 3.2. Given the permissible combinations of perceived body-size, (i.e., large-small, equal, and small-large,) no interactions could be studied, only the main effect of body-size per se. This was not significant: $F(2/37) = 0.0942$ $p = 0.9103$ (see Appendix 3, Table 1). Neither was the interaction between body-size and attribution significant: $F(6/111) = 1.0390$ $p = 0.4040$. Body-size had no significant effect on the attributions to any of the four possible causes: $F(2,37) = 0.1028$ $p = 0.9026$ on attributions made to external forces; $F(2,37) = 1.4170$ $p = 0.2553$ on attributions made to internal forces; $F(2,37) = 0.4266$ $p = 0.6559$ on attributions made to the situation; $F(2,37) = 0.5243$ $p = 0.5963$ on attributions made to "other" forces - see Appendix 3, Tables 2 - 5. Although perceived body-size did affect the perceived nature of ambiguous actions, subjects did not seem to see this as a direct consequence of the actions of different motivational factors on actors with different body-sizes. The more negative perceived nature of the actions of the perceived larger actor were not seen to be caused, as in Duncan's experiment, by stable personality traits as opposed to changeable external environment.

FIGURE 3.1 - PERCEIVED PAIRINGS x MEAN PERCEIVED INTENSITY OF ACTION

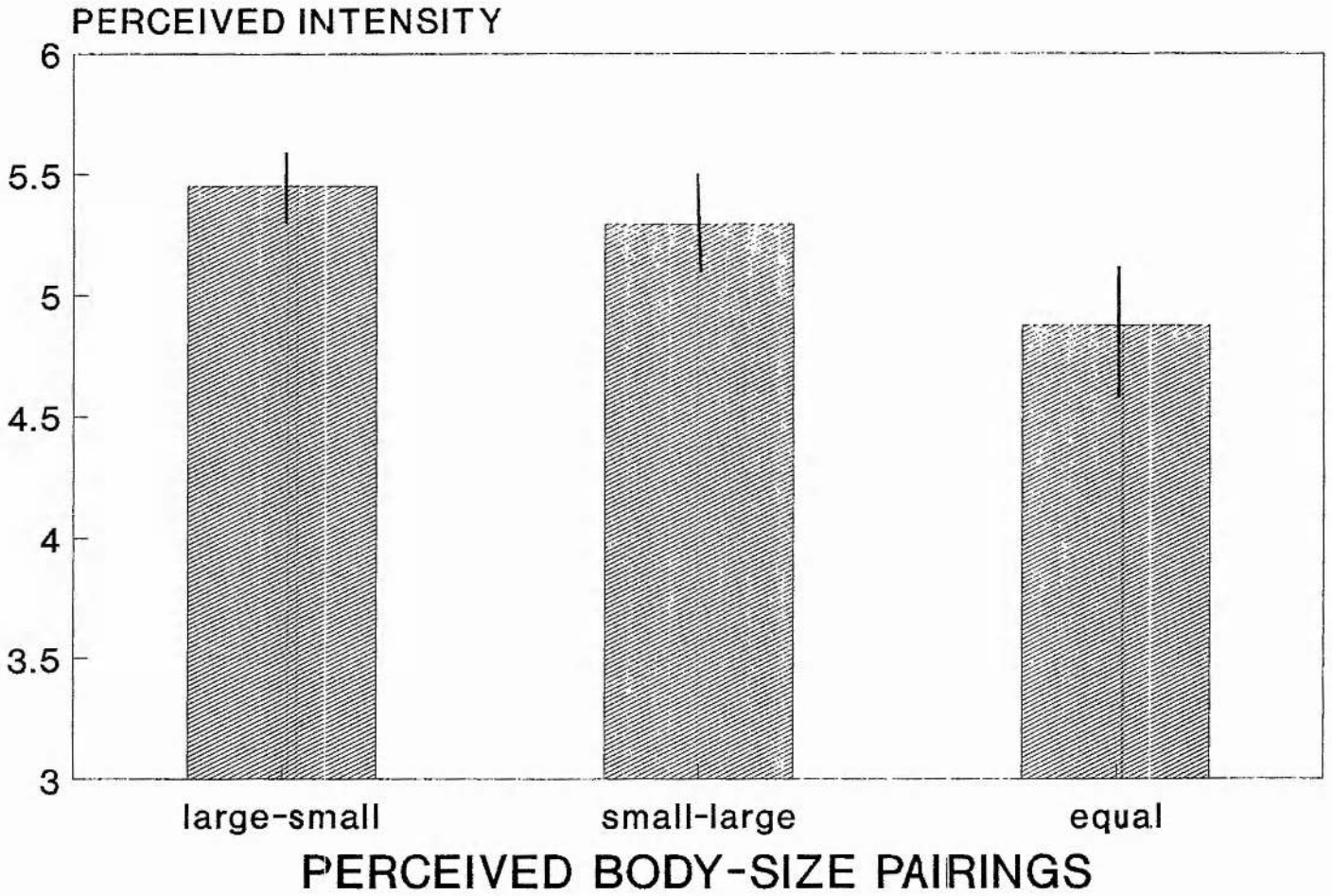
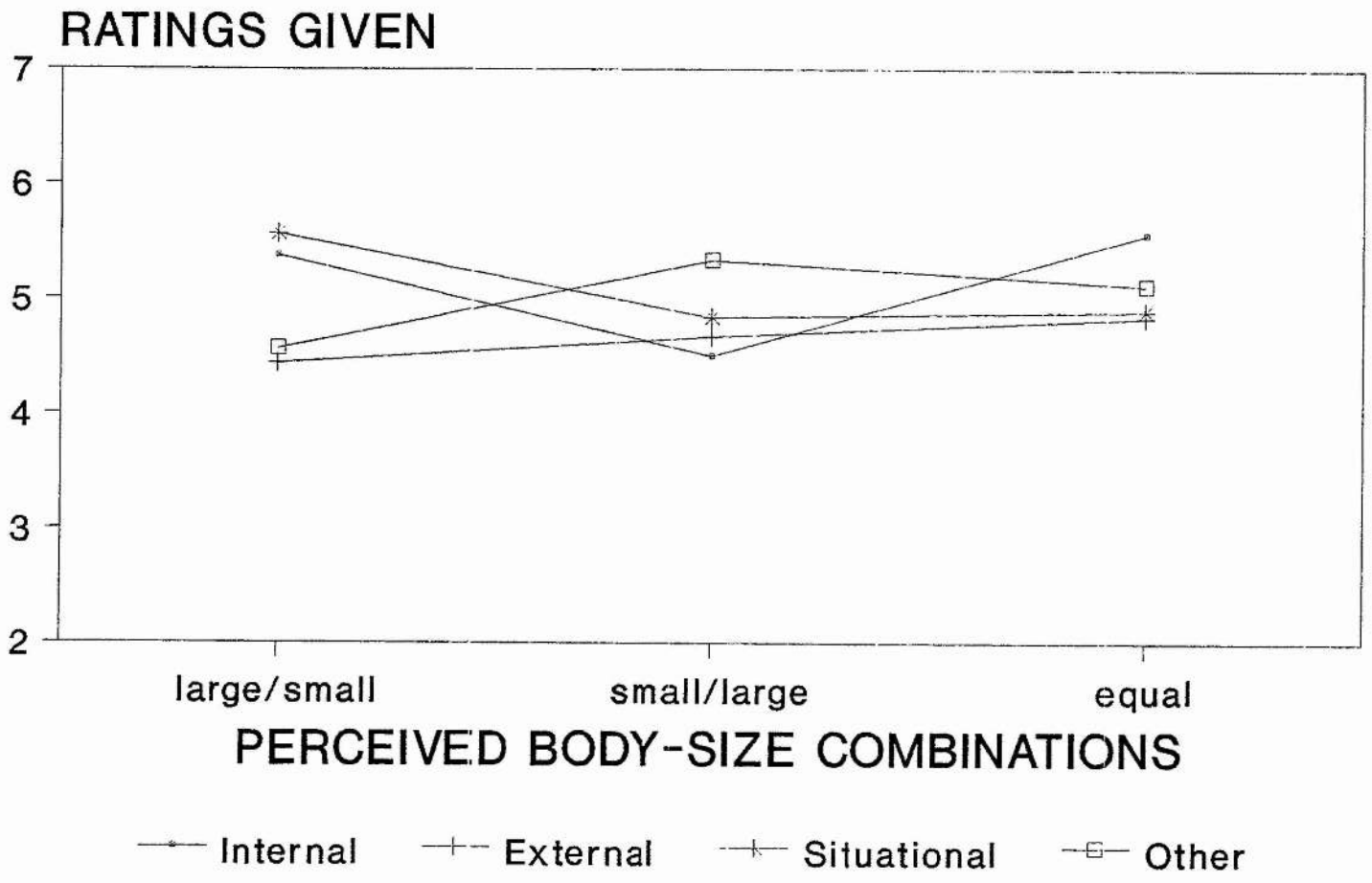


FIGURE 3.2 - ATTRIBUTIONS x PERCEIVED BODY-SIZE OF TARGETS



Any explanation of such an effect must necessarily be tentative and post hoc; however it is conceivable that, body-size being a less immediately obvious variable than race, subjects failed to "think through" their judgments; simply making judgments in a more or less visceral or automatic manner, without trying to attribute causes to them.

Actual body-size: Raw data by actual body-size are presented in Table 3.2. This condition is less relevant to the eyewitness situation, but is included as a check on the data for perceived body-size.

Chi-square was used to determine whether the body-size of the targets in the video-clips affected the number of "violent" ratings given. This was not found to be the case: $\chi^2(1) = 0.4000$ $p > 0.05$. The number of "violent" ratings given was not affected by the body-size of the targets. Table 3.2 shows however that a large proportion of all the ratings given fell into the "aggressive" category.

Considering the "aggressive" category alone, $\chi^2(1) = 0.8043$ $p > 0.05$. Combining the incidence of "aggressive" and "violent" ratings, a similar negative result was found: $\chi^2(3) = 0.5700$ $p > 0.05$. Therefore, although incidence of use of the aggressive/violent categories was found to vary with perceived body-size, it was not found to vary with actual body-size.

Major Category	Large-large	Small-small	Large-small	Small-large
Playing Around	0	1	0	1
Dramatizes	3	1	3	1
Aggressive Behaviour	8	5	5	5
Violent Behaviour	2	3	2	3

TABLE 3.2 - MAJOR CATEGORY FREQUENCIES BY PERPETRATOR/VICTIM BODY- SIZE PAIRINGS

Hence the experimental manipulation of body-size does not seem to have accorded precisely with subjects' perceptions of body-size. Neither did actual body-size have any significant effect on the incidence of "playing around" or "dramatizes" ratings given: $\chi^2(3) = 3.0480$ $p > 0.05$. Hence, actual body-size appears not to affect the number of "negative" or "positive" judgments made.

The effects of the different combinations of body-size on behaviour ratings category used were subjected to chi-square. For the small protagonist/large victim condition, $\chi^2(3) = 4.4000$ $p > 0.05$. For the large protagonist/small victim condition, $\chi^2(3) = 5.2000$ $p > 0.05$. For the small protagonist/small victim condition, $\chi^2(3) = 4.4000$ $p > 0.05$. For the large

protagonist/large victim condition, however, there was a significant effect of behaviour rating category used: $\chi^2(3) = 18.3250$ $p < 0.001$.

The next area to be considered was the perceived intensity of the behaviour of the targets. The raw data are presented in histogram form in Figure 3.3. No effect of actual body-size was found here, $F(3/36) = 1.5223$ $p = 0.2253$ (Appendix 2, Table 2).

Attributional Behaviour: The raw data are presented in graph form in Figure 3.4.

Neither size of protagonist nor victim had a significant effect on attributions of behaviour to external forces: $F(1,18) = 0.1427$ and 0.0857 $p = 0.7101$ and 0.7730 respectively; there was no significant interaction, $F(1,18) = 0.4623$ $p = 0.5052$ (Appendix 4, Table 1).

Attributions to internal factors were equally unaffected by actual body-size: $F(1,18) = 0.2312$ $p = 0.6364$ for body-size of protagonist; $F(1,18) = 1.4015$ $p = 0.2519$ for body-size of victim; $F(1,18) = 1.5632$ $p = 0.2272$ for the interaction (Appendix 4, Table 2).

The same situation held for attributions to the situation: $F(1,18) = 0.4706$ $p = 0.5014$ for body-size of protagonist; $F(1,18) = 0.5145$ $p = 0.4824$ for body-size of victim; and there was no significant interaction between the two, $F(1,18) = 0.0058$ $p = 0.9401$ (see Appendix 4, Table 3).

FIGURE 3.3 - ACTUAL BODY-SIZE PAIRINGS x MEAN PERCEIVED INTENSITY OF ACTIONS

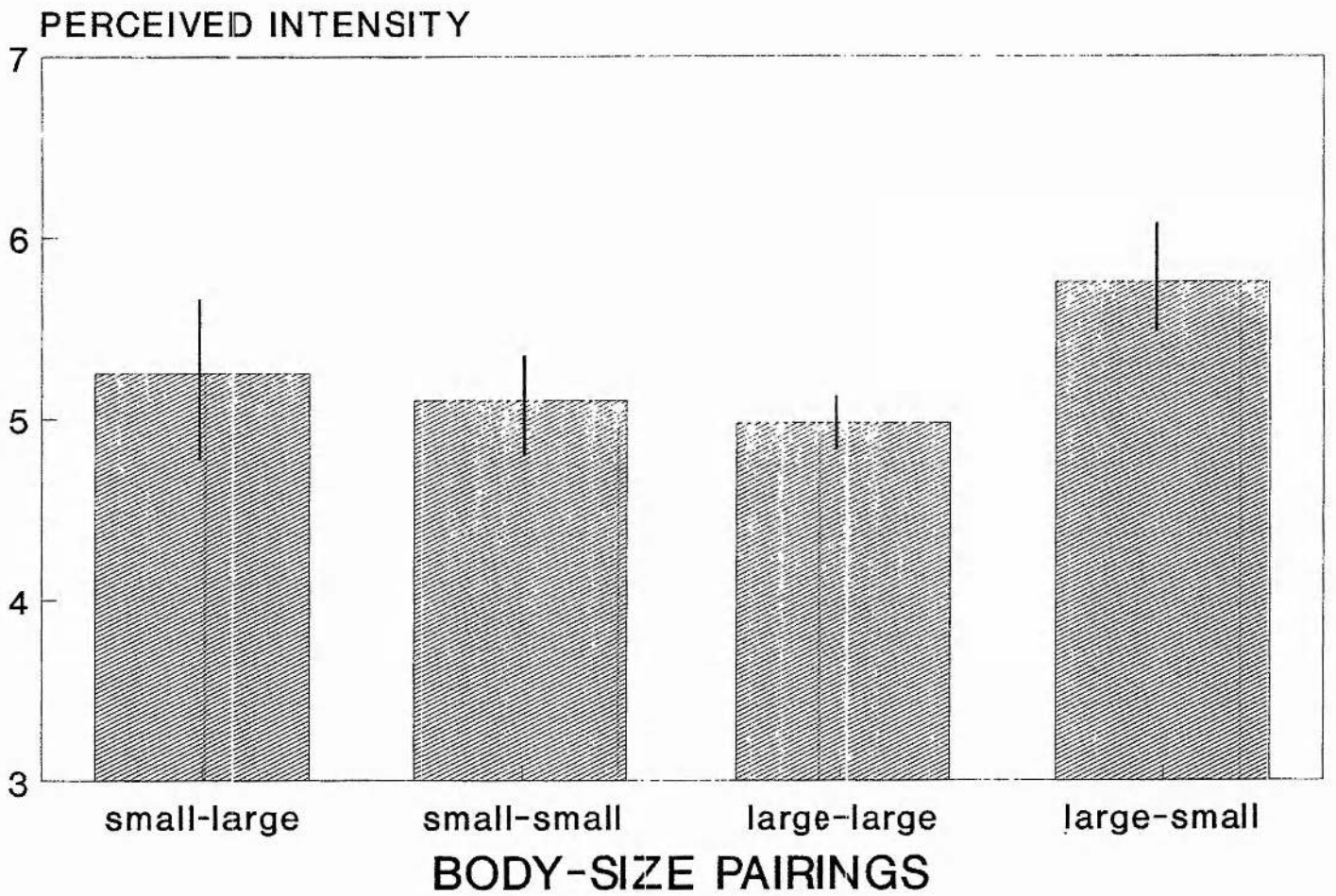
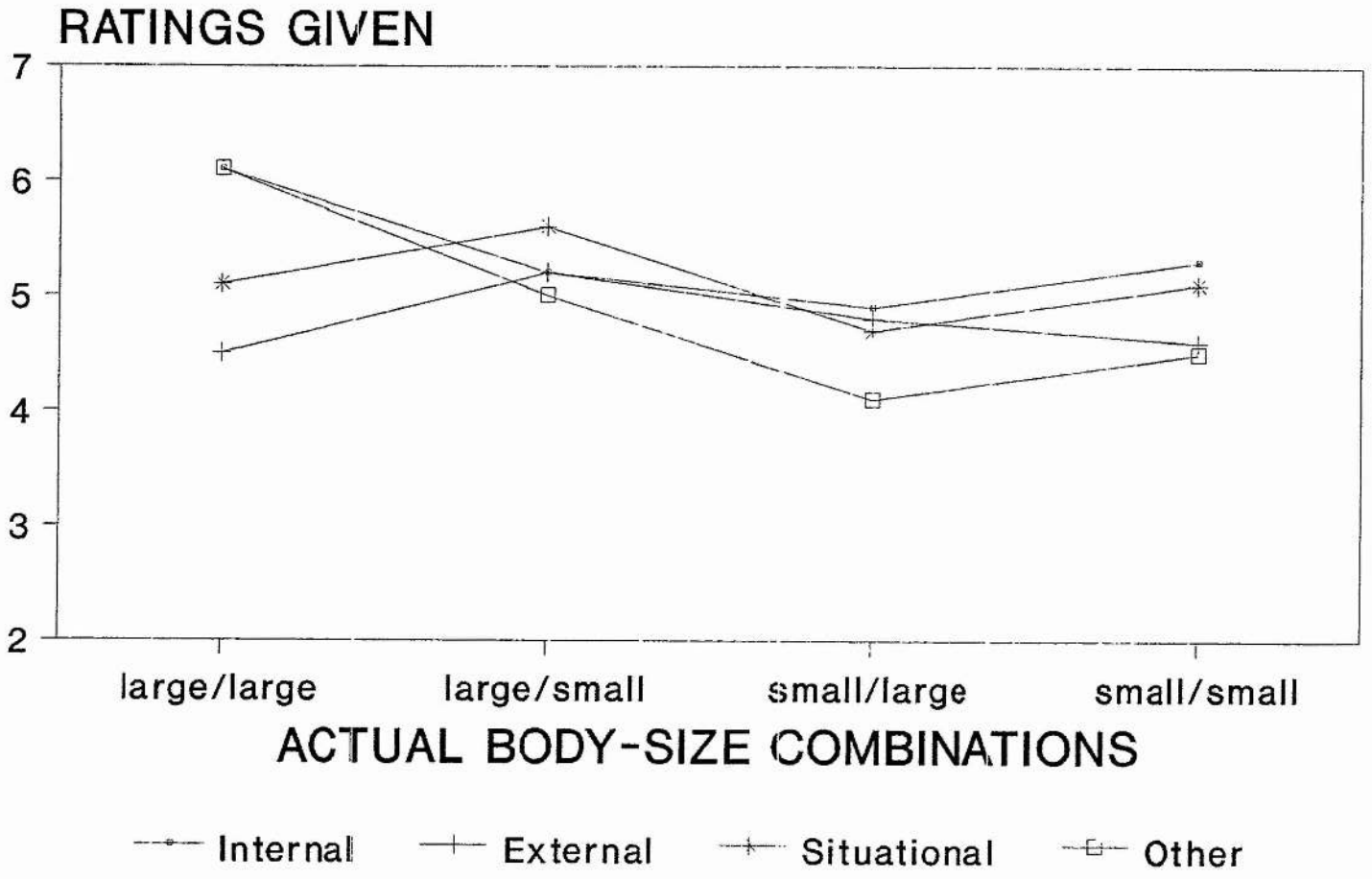


FIGURE 3.4 - ATTRIBUTIONS MADE x ACTUAL BODY-SIZE OF TARGETS



Finally, the number of attributions to "other" factors was not affected by the body-size of the protagonist: $F(1,18) = 0.4353$ $p = 0.5177$, nor was there a significant interaction between body-size of the protagonist and victim and number of such attributions made: $F(1,18) = 1.9990$ $p = 0.1745$ (Appendix 4, Table 4). However, body-size of victim did have a significant effect on number of attributions made to "other" factors: $F(1,18) = 5.9024$ $p = 0.0258$. More attributions to "other" factors were made when the victim was large than when the victim was small: however this is a very difficult effect to interpret.

Discussion

The predictions were that the actions of targets with larger perceived body-sizes would be judged in a more "negative" manner than those of targets with smaller perceived body-sizes, and that these actions would be seen as more internally-driven and more intense than would be the case with a smaller perceived protagonist. The first hypothesis was supported. When the protagonist was perceived as larger than his victim, all the ratings made were found to fall at the negative end of the spectrum. No significant effect was found for the number of "aggressive" ratings given in this condition; although there was a trend in the predicted direction. A significant effect was found for the number of violent ratings, and for the combined total of aggressive and violent ratings given

(i.e., those which fell at the more "negative" end of the spectrum).

This result appears to reflect the powerful action of stereotypes: the actions of persons with a larger perceived body-size seem to be judged more negatively than those of persons with a smaller perceived body-size. It is however difficult to pinpoint the location of this effect. This experiment fails to discriminate between three possible locations for the effect: body-size stereotypes may affect the number of "violent" judgments given, the number of "aggressive" judgments given, or may affect both. In Duncan's (1976) experiment, the race variable acted to increase the number of "violent" ratings given, rather than the number of "aggressive" ratings given. Extrapolating from this result, it might be argued that larger perceived body-size would act in the present experiment to increase the number of "violent" rather than "aggressive" ratings given. However, as mentioned earlier, Duncan failed to investigate the effect of race on the number of "aggressive" ratings given. Additionally, it was hypothesized that, in the present experiment, larger perceived body-size would be associated with a greater number of aggressive judgments, since a correlation between perceived body-size and rated aggressiveness had previously been discovered. No effect of perceived body-size on number of aggressive judgments was however found; although there was a significant effect of perceived body-size on the combined number of aggressive and

violent ratings given. Hence the present experiment does not show whether body-build stereotypes act to increase only the number of violent judgments given, or whether they also have a (smaller) effect on the number of aggressive judgments given.

One possible explanation presents itself. It may be an extremely difficult task to make the distinction between aggressive and violent behaviour. Conceivably, no two subjects are likely to make this fine-grained distinction in exactly the same way. Any two given people may have different ideas about exactly when an "aggressive" action becomes a "violent" action (or vice versa). It could therefore be argued that a combination of these two "negative" categories provides the best measure of the exact operation of stereotypes in this situation.

Another finding that ran counter to hypothesis was that an increase in targets' perceived body-size did not give a corresponding increase in rated intensity of their actions. This hypothesis was however by no means central to the experiment: the main hypotheses could easily have held true despite this. It is possible that the effect of intensity of action in the perceived body-size condition was swamped by the larger effect under consideration. The fact that no effect was found should not necessarily be considered hard evidence that the effect does not obtain: simply that an independent experiment may be necessary to discover this for certain.

An hypothesis that was central to the experiment was that the behaviour of large perceived protagonists would be seen to be due to relatively stable internal traits rather than fluctuating external events. This hypothesis is central to the concept of stereotyping: a personality stereotype is an attribution about enduring traits of the target in question, not external forces acting upon the target.

As suggested above, it is possible that the subtle nature of the stereotypes under investigation here is responsible for the fact that no significant effect was found for the attributions given. Subjects may have failed to "think through" their attributions in a way in which they might have done had a more obvious variable (like race) been employed. The only difference found in attribution of behaviour was a significant effect of the actual body-size of the victim on the number of attributions made to "other" (as opposed to "internal", "external", or "situational") factors. As argued above, the actual body-size of the targets was less important to this experiment than their perceived body-size; hence this is by no means a central finding. In the absence of any other effect on attribution, this is hard to explain; indeed, any explanation that could be given must be of a post-hoc nature.

Perhaps unsurprisingly, the actual body-size of the actors involved had little effect upon interpretation of their actions. This experiment was an investigation into the action of stereotypes, and

these stereotypes were present only in the eyes of the beholders. In a real-world situation, these results suggest that a crime may be interpreted as more aggressive when committed by a larger person than when committed by a smaller person. However this will only hold where the criminal is in fact perceived as being large. Witnesses' perceptions of criminals' body-size appear unlikely always to accord exactly with the actual body-size of those criminals (Clifford and Richards, 1977; Flin and Shepherd 1986).

Against the conclusions reached here it could be argued that the effects seen in the perceived body-size condition were due not to perception of body-size changing perception of action, but to the opposite process. Perhaps when a more aggressive/violent action is perceived, then a larger body-size is deduced for the protagonist. However, this is unlikely to be the case: this argument carries the implication that there was a systematic difference in quality of actions across video-clips. That this was not the case can be seen from the fact that independent subjects rated there as being no significant differences between clips, and the fact that no actual body-size pairing (i.e., no clip) was rated differently from any other. Hence the effects discovered appear to be caused by different stereotyping of targets with different body-sizes.

In a final summary of the results, then, it can be said that while the judgment of ambiguous actions does not appear to be affected by targets' actual body-

sizes, it does appear to be affected by their perceived body-sizes. In the perceived large protagonist/small victim condition, an "ambiguous shove" is seen in a more negative light than in the perceived small protagonist/large victim condition, or when both are perceived to be of equal size. However a marginally significant effect was discovered in the small perceived protagonist/large perceived victim condition for aggressive/violent behaviour: hence it could be argued that the effects discovered are due less to the relative largeness of the protagonist in one condition, as to the perceived body-size disparity in each condition. This appears unlikely however, given that a positive correlation was discovered between perceived body-size and perceived aggressiveness, and that a significant effect was found in the large perceived protagonist/small perceived victim condition.

Finally, this experiment successfully demonstrated a powerful effect of the body-size stereotype on interpretation of an ambiguous action. Where a target is perceived to have a larger body-size, his actions appear to be judged as more aggressive/violent than when the target is perceived to have a smaller body-size. The fact that this effect held even under the conditions reported here - where an actual height difference of just three inches was employed - points to the likely pervasiveness of the effect in real-life conditions, in which larger body-size discrepencies may frequently be encountered. Where real-life

protagonists are larger than those used here, the effect is likely to be even more pronounced.

CHAPTER 4: BUILDING OFFENDER PROFILES: THE EFFECT OF STEREOTYPIC INFORMATION ON THE RECONSTRUCTION AND JUDGMENT OF FACIAL STIMULI

"I am a suspicious-looking person, I know, because I have no chin."
Ian McEwan: Butterflies

Introduction

Experiment 1 considered the effect of a stereotyped feature (body-size) on the judgment of an ambiguous action. Experiments 2 and 3 consider an effect which may be more widespread in the real world: how the presentation of stereotypic information affects reconstruction of facial stimuli (in Experiment 2) and body stimuli (in Experiment 3). Similar paradigms are used in these two experiments, in an effort to allow parallels to be drawn between the two types of stimuli. Facial stimuli are considered first. There follows a review of the literature on facial stereotypes. A number of areas are considered, including the validity or otherwise of facial stereotypes; the way in which these stereotypes may be maintained; memory for facial features; and the stereotyping of faces for intelligence, attractiveness, and occupation. Finally, some methodological issues are considered.

Literature Review

Memory for faces has received more attention in the psychological literature than perhaps any other aspect of the eyewitness situation. Recent years have seen the publication of a large number of books concerned with perception and memory for faces: e.g., Bruce

(1988/91); Davies, Ellis and Shepherd (1981); Ellis, Jeeves, Newcombe and Young (1986); and Young and Ellis (1989). Most of this work does not concern itself exclusively with the eyewitness situation; however, the majority is at least tangentially relevant. One would search in vain for similar reviews concerned with whole-body features, say, or gait. Cutler, Penrod and Martens (1987) argue that, in the "average" eyewitnessing situation, the face and hair are such important memorial cues that they "outshine" all other features.

The literature on perception, judgment, and memory for face stimuli is too vast and wide-ranging to receive adequate attention here, or, indeed, probably in any single volume. The majority of such work is of course outside the present focus anyway. For present purposes, only the influence of stereotypes is considered. This literature is itself extremely large; effects having been shown for perception (e.g., Secord, Bevan and Katz, 1956), memory (e.g., Bartlett, 1932), and judgment (e.g., Macrae and Shepherd, 1989a).

The ubiquitous nature of physiognomic stereotypes can be appreciated if the briefest thought is given to the number of arenas in which they are encountered. Folklore holds that people "with their eyes too close together" are untrustworthy, and that "women with long warty noses may be on intimate terms with the powers of darkness". Writers of fiction employ the stereotypes in a manner that verges on the subliminal:

yet thriller writers, for example, rarely chronicle the adventures of cross-eyed private eyes. Everyday encounters are certainly mediated by such stereotypes, although it is anyone's guess to exactly how great an extent. As Shepherd, Ellis, McMurran and Davies (1978) observe:

> "If systematic distortions in the perception and recall of faces do occur, there may be practical implications for groups such as the police who often collect descriptions of suspects from witnesses to crimes. Clearly, these witnesses, and to an even greater extent the victims, may be in a state of high emotional arousal, and may hold strong attitudes towards the suspect they are describing. In such cases, distortions in recall would be expected to occur." (Shepherd, Ellis, McMurran and Davies, 1978, pp 263 - 264.)

It is this important area which is addressed in the present experiment.

The Validity of Physiognomic Stereotypes: The systematic study of this area has a history which extends at least as far back as Lombroso (1911, cited in Shepherd, 1989): and the history of conjecture about the area may extend centuries further than that. The heyday of research came in the 1920s and 30s, as part of the then-current interest in "constitutional-psychology".

There was some theoretical justification for the view that character could be inferred from facial features. Darwin (1904) pointed out that, since the facial muscles are implicated in the expression of emotion, character might be inferred from measurment of the relative development of these muscles in the individual. And Allport (1937) remarked that

individuals possessed of stereotypically-linked physiognomy might adopt traits imputed to them by the rest of society. However, the findings were mostly negative (see, e.g., Cleeton and Knight, 1924; Hull, 1928; also Atherton and Ley, 1971). Hence the area was neglected for many years, with the exception of work by Sheldon and his colleagues (see Chapter 3).

Relatively recently, however, three papers have appeared which seem to be minor exceptions to the generally negative rule. Squier and Mew (1981) studied the correlation between facial structure and score on the 16PF scale in groups of subjects with differing faces. They found long, angular faces to be associated with naïvete and naturalness; and short, square faces with shrewdness, discreetness, and susceptibility to emotion. Two other relatively recent studies with positive results were reported by Terry and Snyder (1972), and Terry (1975), who found that high-school students showed not only highly consensual but also accurate judgments in identifying three different social categories from photographs of unknown students. Since one of the categories investigated was "beauty queens", however, one must speculate as to the difficulty of this task.

Taken alone, these studies may not constitute undeniable evidence that physiognomic stereotypes are, after all, valid: but certainly imply that a second look at the area might be valuable.

Mechanisms Underlying Physiognomic Stereotypes: With the general failure to discover any relationship between physiognomy and psyche, the psychological literature turned to descriptive studies of the stereotypes in question (see, e.g., Bradshaw, 1969; Hochberg and Galper, 1974; Hurwitz, Wiggins and Jones, 1975; and MacGregor, Abel, Bryt, Lauer and Weissmann, 1953), and, more ambitiously, attempts to discover how and why such stereotypes arose. Secord (1958/59; Secord and Bevan, 1956; Secord, Bevan and Katz, 1956; Secord, Dukes and Bevan, 1954) was the founding father of this research. His methodology has since been advanced by Cunningham (1986) and Berry and MacArthur (1986), whose work lies outside the scope of the present review.

Secord (1958) suggested two mechanisms by which stereotypic responses to faces might be mediated. First, it is (virtually) always a salient cognitive response to another person to assign him/her to certain social groups (by, e.g., age or sex), and this categorisation can be performed on the basis of certain undoubtedly valid physiognomic cues. Such categorisation will provide a basis for stereotypic attributions. Second, stable dispositions may be inferred from relatively transient facial expressions: e.g., inferring on the basis of a passing smile that a target person is of a happy disposition. Attribution theory holds that, in the absence of other dispositional information, such cues are likely to be utilised by the perceiver. A third possible mechanism

is mentioned by Shepherd (1989): It is possible that certain facial attributes have evolved as means of communicating such selectively-important states as dominance or dependence, and these attributes form the foundation of many social judgments. These remain the cues most likely to underlie stereotypic phenomena associated with physiognomic information (Shepherd, 1989).

Memory, Judgment, and Attributes of the Face: The present sub-section briefly considers the literature on face-memory, and attempts to demonstrate that, while much previous work has considered individual features of target faces, emphasis should now be shifted to take more account of global attributes. Specifically it is argued that, despite the wealth of research on face-memory in the psychological literature, little research has addressed the issues most important in the social situations in which face-memory is usually employed in the real world. Research into the effect of stereotypes on face-memory is well-positioned to help fill this gap. The sub-section ends with a brief consideration of the way in which judgments of facial stimuli are made, contending that similar arguments can be made for this paradigm.

The earliest work on face memory (which used recognition paradigms almost exclusively: recall paradigms having been used with regularity only during the last fifteen years or so) was concerned with the relative memorability of different parts of the face

(e.g., Hanawalt, 1944; Howells, 1938). This work generally supports two complementary conclusions: that, with the exception of the hair outline, the eyes and mouth are the most salient parts of the face for memory (presumably because they are "expression-carrying" - Clifford and Bull, 1978, p 72, and Argyle, 1967 - and also vital to social interaction - Argyle and Cook, 1975), and that upper features tend to be more important than lower features. A good example is the work of McKelvie (1976), who presented subjects with photographs of faces with or without eyes or mouths masked, and then tested recognition memory for these faces. He found that masking eyes caused significantly more errors than masking mouths. Goldstein and Mackenberg (1966), testing a population of school-children, found that covering upper features had a greater effect on subsequent recognition of school friends than covering lower features; and Friedman, Reed and Carterette (1971) found facial features to be important to recognition in the order: nose, eyes, forehead, mouth. Ellis, Davies and Shepherd (1975), using a Photofit kit, found that subjects selected features in the order: hair and forehead; eyes; nose; mouth and chin. Shepherd, Ellis and Davies (1977) had subjects write descriptions of ten of one hundred target photographs of male faces. From four hundred descriptions generated, frequency counts were made of the number of times each feature was mentioned and the number of subjects using the particular feature description. Of thirteen discrete

facial features identified, upper face features received most attention: Hair, forehead, eyebrows, and eyes together accounted for nearly half the total number of feature descriptions given. A similar result was found by Ellis, Shepherd and Davies (1980), using colour prints of two male faces, with varying delay intervals. The pattern of results achieved was strikingly close to that reported by Shepherd *et al*. The primary importance of the hair is stressed in a literature review by Shepherd, Davies and Ellis (1981).

For further research on the relative importance of different facial features to memory, see, e.g., Garneau (1973); Langdell (1978); Nash (1969); Seamon, Stolz, Bass and Chatinover (1978), and studies of eye-movement during face perception by Cook (1978); Luria and Strauss (1978); Walker-Smith, Gale and Findlay (1977); and Yarbus (1967).

While these findings have great value to anyone conducting research into face memory, more recent work has emphasised the importance of the facial Gestalt over individual features. This has an intuitive appeal when everyday experience is taken into account: e.g., the fact that it is possible to recognise as adults people we knew only as children, by which time individual features will presumably have changed, leaving only the pattern the same.

Perhaps the most compelling evidence for this argument comes from Fisher and Cox (1975). They revealed to the observer progressively greater amounts

of target photographs of famous persons, doing so either horizontally or vertically. This allowed them to see not only which features were most important in facial recognition, but whether pairs of features had additive or multiplicative effects. They found that lower features contributed 15% more information in the context of upper features than when presented alone, and that the addition of eyes produced the greatest increase in recognition performance. However, only 12% of subjects were able to recognise target faces when eyes were presented alone. Further evidence comes from Homa, Haver and Schwartz (1976), who found that features presented within the context of the face were significantly better remembered than features presented alone.

Others have found face-memory to depend upon attributes of a whole face: e.g., Cross, Cross and Daly (1971) found memorability to be an inverse function of normality: i.e., the more normal a face, the lower its memorability (see also below); and Peters (1917, cited in Ellis, 1975) found memorability to be a function of "pleasantness". A similar finding was made by Shepherd and Ellis (1973) for the attractiveness of a face: memory of an undistinctive target face was at chance level after only five weeks. However they found no evidence for increased forgetting with time for a distinctive face. Clifford and Bull (1978) make the telling observation that, when using Photofit, subjects rarely describe targets in terms of discrete features, using instead such

global descriptions as "fierce-looking". This is supported by Mueller and Thompson (1986, Expt. 2). They had subjects generate traits applicable to twelve famous personalities, either before or after a recognition task which involved picking from an array famous faces which had previously been presented. Subjects who wrote down trait information before the memory task were expected to analyse the presented faces in terms of "deep traits"; the others in terms of "shallow traits". As predicted, subjects who wrote the largest number of traits ("high-trait-generators") made fewer false alarms on testing than "low-trait-generators". Mueller and Thompson termed the process of trait-generation "stereotyping". Although their use of the term differs from that of most social psychologists, their results indicate that subjects did not rely upon feature information to improve performance in the memory task, but on some supervenient quality.

All of this is evidence that future work on the memorability of face stimuli would do well to concentrate on attributes of faces rather than discrete features. Shepherd, Davies and Ellis (1981) conclude their review of face fragmentation studies with the following comment:

> "Upper features are generally better cues for reintegrative purposes than lower ones. Eyes in particular emerge as powerful aids, though eyes in isolation are not particularly effective. This latter finding underlines one deficiency of all fragmentation studies, namely the way in which the normal facial Gestalt must be broken up, giving rise, perhaps, to abnormal or unusual processing strategies. Before drawing any more general conclusions regarding the saliency of

> various aspects of the face, studies which have striven to retain a normal whole-face at all stages of study should be surveyed."
> (Shepherd, Davies and Ellis, 1981, p 119)

A similar argument can be made for stimuli mediating judgments of target faces. To take just one example, judgments of intelligence do not appear to be made on the basis of any single facial cue. Gurnee (1924) found that no single facial part correlated significantly with whole-face judgments of intelligence; the balance of parts appearing essential. Secord *et al* (1954) found no single physical feature of a face to be associated with intelligence; although Laser and Mathie (1981) did find correlations of intelligence ratings with facial length and eyebrow thickness, and McArthur and Apatow (1983 - 84) found a correlation with eye size. Shepherd (1989) concludes his survey of this literature with the comment that:

> "It is unlikely that any single trait of physiognomy will be related to judgments of intelligence. What evidence there is points to a constellation of feature attributes which add up to general attractiveness."
> (Shepherd, 1989, pp 300 - 301.)

This conclusion, reinforced by the research of Brunswick (1956), Clifford and Walster (1973), Cook (1939), and McArthur and Apatow (1983 - 84), further strengthens the notion that global features of faces are those upon which social judgments most depend. It is a well-documented phenomenon that the more "average" or "normal" the inter-relationship of facial features, the more likely is a face to be judged attractive (Benson and Perrett, 1991/2; Pollard 1986,

cited in Shepherd, 1986). It appears important for research in this area to take more account of the attributes of facial stimuli (attractiveness, stereotypicality, etc.) than has been the case in the past.

It is to this literature that the present study aims to contribute. Particularly germaine to Experiment 2 is previous research considering memory for and judgment of physiognomic stereotypes associated with intelligence, attractiveness, and criminal occupations in facial stimuli. These are reviewed briefly below. Obviously, studies involving stereotypes of honesty and aggressiveness would also be applicable here: however I am not aware of any work in these areas, other than that by Macrae and Shepherd (1989).

Intelligence:

> "The assessment of intelligence from the face has preoccupied psychologists and physiognomists for centuries. Before the advent of intelligence tests, and after, physiognomists tried to devise an index based upon measurements of the skull and face which would yield a measure of intellectual capacity. Lavater (1804) in the eighteenth century believed he had found such a formula, though it was never systematically tested." (Shepherd, 1989, p 299.)

The majority of research into the physiognomic stereotypes of intelligence was carried out during the "constitutional psychology" boom of the mid-century, and most was concerned with testing the validity of such stereotypes. Here, constitutional psychology had an unusual, if modest, success. Gaskill, Fenton and Porter (1927) had judges rank for intelligence

photographs of 11 - 12 year old boys. The correlation between these ranks and measured IQ was .42: although the very extreme range of IQ scores used renders the generalisability of the results questionable. Anderson (1921), Cook (1939), Laird and Remmers (1924), and Moriwaki (1929) all found median positive correlations up to .3.

Shepherd (1989) summarises this literature as demonstrating that judges usually agree on rankings of target faces, but that such orderings are usually only weakly correlated with external measures. Work by Hollingworth (1935 - who compared the attractiveness ratings given to gifted adolescents with those given to adolescents of normal intelligence,) and that cited above by Clifford and Walster (1973) and McArthur and Apatow (1983 - 84), makes Shepherd's conclusion appear all the more valid. This is supported in part by a finding of Shepherd, Ellis, McMurran and Davies (1978). Photofits of a target face were constructed under the belief that the target was either a murderer or a lifeboat captain. When the target was believed to be a murderer, generated faces were judged to be significantly less "intelligent" and significantly less "good-looking" than when he was believed to be a lifeboat captain.

Such findings show the necessity of taking into account the "attractiveness" factor when seeking differences in perceived "intelligence" in experimental stimuli, as in the present experiment.

Attractiveness: Research into facial attractiveness has had a somewhat different slant from that concerned with intelligence. Judgments of attractiveness cannot be tested against external criteria; hence the majority of research has considered levels of inter-rater agreement. While impressive research has been conducted here, (see, e.g., Cash 1981, Berscheid and Walster 1974,) this is not the area of concern here.

The stereotypical basis of facial attractiveness has come under psychological scrutiny only relatively recently, perhaps because of the methodological difficulties involved in the work. The main problem is probably in trying to isolate factors that consistently affect perceived attractiveness across the large range of possible stimulus groups (infants to senior citizens of both sexes and all races), all of which presumably boast "more" and "less" attractive members. The most important single study in the area however is probably that conducted by McArthur and Apatow (1983 - 84), in which they discovered attractivess in adult males to be associated particularly with increased maturity of features (the opposite of the case with adult females). A similar discovery was made by Keating (1985).

In attempting to explain such results, the literature has expended much energy in considering the communicative functions of "babyish" and "mature" stimulus configurations (e.g., Cunningham, 1986) - in females, for example, perceived "submissiveness" may have selective importance, while the opposite may hold

for males. Little energy has however been expended on arguing for the idea of a Platonic form (or "optimal stimulus configuration"), to which more attractive faces may tend. Of course, it is possible that the Platonic form tends towards the "mature" in males and the "babyish" in females (Benson and Perrett 1991/2).

Different approaches to the quantification of human facial attractiveness have been taken by orthodontists - who have related measurements of parts of the face to measurements of facial attractiveness - and studies by Hirschberg, Jones and Haggerty (1978) and Milord (1978) have applied multi-dimensional scaling to the same problem. Hirshberg *et al* (1978) used photographs of nine male black and nine male white faces for target stimuli. Thesetargets were known to vary on a previously-developed Face Differential Scale. Subjects were asked to rate the similarity of all possible pairs of stimuli, and also to rate the faces on a number of physical and psychological characteristics. Additionally, physical measurements were taken of a number of facial features of each photograph. The dimension accounting for most variance in the judgments was the race dimension, with which no psychological ratings were correlated. The second dimension (called "desirability") was related to ratings on such scales as "likeable", "attractive", and "pleasant". There were positive correlations with "chin type" and "nose profile". The third dimension was labelled "masculine maturity", which was correlated with the psychological

characteristics of "strength", "masculinity" and "maturity", and ratings of the physical features "moustache", "beard" and "sideburns". The fourth dimension was "face shape", correlated with one physical and one psychological dimension; and the fifth dimension was correlated with ratings of "relaxed" vs. "tense" and "eye-shape". Finally, the sixth dimension was related to face flatness and neck thickness for white, but not for black, subjects. As Shepherd, Ellis and Davies (1981) point out, it is an interesting feature of these results that in some cases dimensions were related to global aspects of the target faces but not to specific features; while in other cases they were associated with a specific physical feature, but no psychological characteristic. In a similar experiment, Milord asked subjects to judge the similarity of pairs of young white male target faces, and found that the dimension that accounted for most variance correlated positively with ratings ofthe "pleasingness" and "beauty" of the faces, and negatively with ratings of their "uniqueness".

Some studies appear to show an advantage of attractive over unattractive faces in memorability; a finding which might be taken to imply a similar advantage for "intelligent" over "unintelligent" faces. Cross, Cross and Daly (1971) found better recognition memory for attractive than unattractive faces, across a range of social categories, in an experiment in which subjects were led to believe that

they were to be judging the facial beauty of a set of stimuli. This was not caused by length of exposure time - since subjects spent longer attending to faces of medium attractiveness - but may have been due to more "active attention" paid to the facial Gestalt (this interpretation can however be queried: see Shepherd, 1981). The findings of Benson and Perrett that more average faces are perceived as more attractive throw some doubt over this assertion, since a number of studies in the literature appear to have found poorer memory performance for typical than atypical faces (see Valentine, 1991, for a brief review). Yarmey (1975) found that faces rated *low* on attractiveness were later better recognised than faces rated *high* or *medium*. Memory for faces has consistently been shown to be better following more elaborate processing (see, e.g., Bower and Karlin 1974, Winograd 1978). Winograd (1981) holds that this effect is due to the greater chance of encoding a distinctive facial feature following more elaborate encoding.

Although there does appear to be a high level of cross-cultural agreement on standards of facial attractiveness (Cross and Cross, 1971; Thakerar and Iwawaki, 1979), these levels may not always be so high as those achieved within a single culture (compare, for example, the inter-rater correlation achieved by Madden and Hollingworth, 1932, with that achieved by Dion, 1973; and see also Milord, 1976); hence it is probably practical to limit target stimuli in face-

processing experiments to examples drawn from just one culture, at least until such time as these differences have been quantified.

Occupations: Stereotyped notions of both criminal and conventional occupations have received attention in the literature. It is the former which are of greater interest here.

Three studies have shown agreement across judges in the assignation of male faces to criminal categories. Shoemaker, South and Lowe (1973) had subjects select from an array of twelve photographs those most and least likely to have committed one of four types of crime. Judges agreed significantly in both positive and negative directions, indicating a shared facial stereotype of four types of "criminal" - homosexual, murderer, robber, and traitor. Bull and Green (1980) presented ten photographs of young male adults to forty-eight members of the public, who had to ascribe each of eleven listed crimes to the faces. No consensual agreements were found for faces ascribed to the crimes of arson, theft, rape, or burglary. However, consensual agreements were found for the faces that best fitted the crimes mugging, robbery with violence, company fraud, soliciting, possession, and gross indecency. In each of these cases, one face was chosen more frequently than the others. Golstein, Chance and Gilbert (1984) arranged facial photographs of white middle-aged men into five equal, separate arrays. Subjects were told that each array contained

one mass murderer, one armed robber, one rapist, one medical doctor, one clergyman, and one engineer. Attributions of occupation category to photographs was found to be significantly non-random. A small number of photographs was frequently selected for what Goldstein *et al* term "good guys" and a small number for "bad guys". These choices also tended to be occupation-specific, although this effect was more pronounced for criminal than non-criminal occupations. This effect may be caused by "some kind of unintentional 'matching-to-sample' technique, wherein the 'sample' is a prototypic image or other memory representation of a face" (p 551).

It is unfortunate that none of these studies used target persons with known criminal records, since no inference can be made about the accuracy of these inter-rater agreements. Lombroso (1911) would argue that the agreements might well be veridical, holding that there is indeed a set of physiognomic signs of criminality, including low forehead and facial asymmetry.

This very issue was addressed by Thornton (1939). He selected case records and photographs of twenty criminals, showing the photographs to a group of subjects who were required to say which of four possible crimes they thought the criminal had committed. Although the accuracy of these judgments was above chance, Thornton concluded that the effect was so slight as to be negligible. Hence it appears unlikely that judgments made in the experiments cited

above were in fact veridical, although the possibility cannot be discounted entirely.

Kozeny (1962, cited in Bull and Green, 1980) obtained photographs of seven hundred and thirty convicted criminals, and divided these into sixteen categories by crime type. From each category a composite portrait was produced, which showed a statistical difference in the physiognomies of the different groups. This effect is perhaps less likely to be caused by genetic factors than by social expectations of the behaviour of persons possessed of certain physiognomies: Cavior, Hayes and Cavior (1975, cited in Bull and Green, 1980) found that low physical attractiveness contributed to criminal deviancy. This effect appears to operate even in the world of very young children: see Rich (1975) and Bull and Green (1980) for further development of this argument.

The real-world importance of such physiognomic stereotypes does not end here. As Bull and Green remark:

> "...witnesses of a criminal incident may unwittingly permit in their identifying of a suspect, their expectations about him (or her) to play a role. That is, for example, when witnesses attempt to pick a person out from an identification parade they may merely pick out the person who best resembles their expectation of what a certain criminal should look like." (p 50)

Three studies in the literature have considered memory and occupational stereotypes (criminal or otherwise). These studies are discussed in the next sub-section, along with the other study which is

especially germaine to the present work, that by Shepherd (1991).

Stereotypes, Memory and Judgment of Face Stimuli: Given the number of papers that have been published on the subject of face memory, it is perhaps surprising that so few have considered the effect - particularly relevant here - of stereotypic information upon face memory. Judging from the powerful effects demonstrated for stereotypes in the literature, (from studies of biographical material by Snyder and Uranowitz, 1978c, to actions performed in video-recordings by Duncan, 1976, to pictures shown to children by Martin and Halverson, 1983,) one would be naïve to expect stereotypic information to have no effect in this instance. The few studies which have addressed this issue are reviewed below.

Klatzky, Martin and Kane (1982) report a study which, although utilising a recognition rather than a recall paradigm, is nevertheless germaine. Klatzky *et al* were studying categorical perception of faces by occupation. Subjects were found to be able to assign faces to occupations at above-chance levels in a forced-choice task. On a subsequent memory task, recognition performance was found to be superior for stereotypically-linked than for non-stereotypically-linked faces: implying that category membership may play a role in the encoding of faces.

Bartlett (1932) demonstrated an analagous effect of occupational stereotypes on recall. He found verbal

recall of previously-presented line-drawings of military personnel to be biased in the direction of then-current stereotypes of soldiers.

Two experiments that are particularly relevant to the present work are those by Macrae and Shepherd (1989a) and Shepherd (1991). Both studies demonstrate the distorting effect of stereotypic information on later reconstruction of facial stimuli, and judgment of those reconstructed stimuli by independent judges.

Macrae and Shepherd (1989a) present what is perhaps the best controlled study yet to be conducted into the effect of stereotypes on memory for faces. In pilot studies, they identified four target photographs of males, two of which had been rated at the opposite ends of the honesty dimension and two at the opposite ends of the aggressiveness dimension. These targets were all matched for age and attractiveness. Each subject viewed a videotape consisting of a picture of one of the faces and a commentary describing either an assault or a theft. Subjects in the "assault" condition were presented with either a face that had previously been rated as high or a face that had previously been rated as low in aggressiveness, while subjects in the "theft" condition were presented with either a face that had previously been rated as high or one that had previously been rated as low in honesty. The commentary informed them that the individual in the photograph was suspected of committing the crime in question, and they were asked to rate the likelihood of the target's guilt, using

only the face information available. Macrae and Shepherd found that, for both incidents, subjects considered the defendant with the stereotypically-linked physiognomy more likely to be guilty.

This study is somewhat reminiscent of a study by Shepherd (1991 - personal communication). Subjects were presented with a male target face along with biographical material supposedly pertaining to the target person. In one condition the target was said to have received a 2i degree, and in the other a 2ii degree; otherwise this material was identical across conditions. Later, subjects were asked to reconstruct the face they had seen, using a Photofit kit. The attractiveness of these reconstructions was then rated by independent judges. Those made by subjects who were told that the target received a 2i degree were judged significantly more attractive than those made by subjects who were told that the target received a 2ii degree. The study was an elegant demonstration of the effect of the stereotypical notion of the link between intelligence and attractiveness on the reconstruction of facial stimului.

Experiment 2 constitutes an attempt to replicate this finding, and also to investigate the effect of the criminal stereotype isolated by Macrae and Shepherd (1989a) on reconstruction of a target face, in an attempt to clarify the results achieved in the previous experiment, and to extend present knowledge of the effects of stereotypic information upon memory and judgment in the eyewitness arena.

Methodological Points:

1) TIME OF PRESENTATION OF STEREOTYPIC INFORMATION

Classically, there are two strands to the research into the effects of stereotypic information on memory or judgment. Investigators have considered the effects of stereotypes presented at encoding of target material and (less frequently) those of stereotypes presented at retrieval. The literature on the effects of stereotypes presented at encoding is a particularly complex one. This literature will be exhaustively reviewed in Chapter 6, but for the moment a more or less random sampling of the results reported might give an idea of its contradictory nature. Rothbart (1981) found that stereotype-consistent information was better recalled than inconsistent or irrelevant information; O'Sullivan and Durso (1984) also found that consistent information was best remembered, but only when very inconsistent information was also presented; Hastie and Kumar (1979) found better recall for inconsistent information; and a similar result was found by Crocker, Hannah and Weber (1983).

The social cognition literature therefore gives no a priori grounds to predict what type of information - consistent or inconsistent with the physiognomic stereotype of the criminal - will be better remembered by an eyewitness when stereotypic information is present at encoding; or indeed whether either will be better remembered. Given that there is potential here to elicit extra information from an eyewitness, as

well as potential for an eyewitness' account of a crime to undergo systematic distortion, it seems only sensible that systematic investigations into the area should be performed.

A smaller literature exists on the effects of stereotypes presented at retrieval: one which centres around the ongoing controversy surrounding the seminal paper by Snyder and Uranowitz (1978c), on the effects of a "lesbian stereotype" on memory for a biography of the fictional character, Betty K. Calls for greater emphasis in the literature on stereotypes presented at retrieval (see, e.g., Bodenhausen and Lichtenstein, 1987; Bodenhausen and Wyer, 1985,) have unfortunately gone all but unanswered; because of either the thorny nature of the area, or the fact that the majority of models of effects of stereotypes on memory take little account of the possibility of this effect (see Chapter 6). Bodenhausen and Lichtenstein write:

> "There is a great deal of research documenting the importance of encoding operations and organizational processes, but there is a relative dearth of evidence provided by social cognition research documenting retrieval effects" (p 879).

Experiment 2 will take the opportunity partially to fill this hole in the literature by considering the effects of stereotypes presented both before and after target information.

2) THE NATURE OF TARGET STIMULI Ekman (1978) isolated three categories of facial information by which stereotyping might be mediated. These are: 1. static

signs with a very slow rate of change, e.g., the structure of the skull, 2. faster-changing signs, e.g., quantity of hair, wrinkling of skin, and 3. rapid signs: like muscular contractions which temporarily alter the shape of a facial feature, or facial expressions which communicate emotion. Most real-life judgment and memory tasks presumably rely more heavily upon the first two types of sign, and less upon the third, since the third type is usually likely to carry least information. Most laboratory stimuli, however, fail to discriminate between the types of sign. It is difficult to tell, for example, whether the sneer in a still photograph is the long-term consequence of a target's sneering disposition, or a short-term reaction to the infacility with which the photographer handles his camera.

The use of schematic faces rather than still photographs as experimental materials has the advantage that rapid signs are eliminated, but the consequent disadvantage that much individuating information is lost. Artist's sketches and Photofit reconstructions may be more realistic; but may produce poor likenesses. Photofit appears to produce a worse likeness than a witness' own sketch (Ellis, Davies and Shepherd, 1978, and see Davies, 1981, for a discussion of different face recall systems). Also, the lines which appear between Photofit features appear to disrupt processing (see, e.g., Ellis, Davies and Shepherd, 1979). Both Photofit and Identikit (Laughery, Duval and Fowler, 1977, cited in Davies,

1981) appear to lack sensitivity. Neither kit appears able to utilise all the information available in a target face. Video-recordings presumably produce the most realistic materials, but have the disadvantage that it is hard to generalise from one very specific set of video-recorded faces to the general population. There is also the problem of selecting lighting conditions, poses, and expressions, and keeping these constant across targets.

Photofit was the system employed by Shepherd, Ellis, MacMurran and Davies (1978), who claimed that, although the system often produced poor likenesses, it did seem a useful method by which to investigate systematic biases in visual recall. The Mac-a-Mug system, while basically similar, is something of an improvement over Photofit in that it eliminates horizontal lines across the face, and requires no special training to use (although it does not overcome the problem of interference caused by subjects' perusing a large number of features: Davies, Shepherd and Ellis, 1979; Laughery, Fessler, Lenorovitz and Yoblick, 1974). The many facilities which Mac-a-Mug offers for enlargement and reduction of facial features, movement of features relative to one another, and application of many different types of shading, also make it an extremely subtle instrument. It might therefore be considered to go some way towards balancing the pay-offs of photographic stimuli and schematic faces. A subject can add or subtract rapid signs, and make the face more or less schematic,

at will. Mac-a-Mug is not the perfect system for this type of research, but is probably the best currently available.

One other change was made here from the Shepherd *et al* methodology. In the original experiment, only female subjects and judges were used, presumably because of the (arguable) female superiority with such systems (see Ellis, 1975; McKelvie, 1978). In the present study, both male and female subjects were used, since this gives a more realistic representation of a genuine eyewitnessing situation. Besides, what gender difference there is, may lie in the opposite direction for violent crimes like assault, and vary with type of crime witnessed (Clifford and Scott, 1978; Kuehn, 1974). Hence it seemed appropriate to use both male and female subjects in the present experiment.

Experiment 2

The two pairs of reconstruction conditions in this experiment are reported separately as Experiments a and b.

Experiment 2a

The hypotheses of this experiment are listed below.

1. Where subjects are told a target committed an assault, the faces generated should be judged more stereotypically "aggressive" than when subjects are told that the target committed a theft. Where subjects are told a target committed a theft, the

faces generated should be judged more stereotypically "dishonest".

2. Effects should be less pronounced, although still significant, when stereotypes are presented after than when they are presented before to-be-remembered information.

Methodology

Subjects: Subjects in this experiment were fifty-four undergraduate and postgraduate volunteers, who took part for a small fee. Thirty subjects were female, twenty-four male; with an age range of 18 - 26 years.

Apparatus: An Apple Macintosh computer loaded with the Mac-a-Mug programme was used for both generation of target images and their later reconstruction.

Materials: Original target stimuli were generated by asking eighteen volunteers (ten male, eight female, all students aged 20 - 25 years) to come separately to the laboratory and use the Mac-a-Mug programme to create face images. They were asked to create images of adult male faces which they knew at least moderately well. Each volunteer was given fifteen minutes for this task. All were explicitly informed of the intended use of these images. The eighteen faces thus generated were submitted to six independent judges (Honours Psychology undergraduates from the University of St Andrews) who rated them on nine-point scales for the dimensions intelligence,

attractiveness, honesty, and aggressiveness. Only target faces that were balanced on these dimensions were used in the experiment. Six target faces fulfilled this criterion (see Appendix 6, Tables 1 - 4).

A two-hundred word biography was constructed in two stages: four randomly-selected postgraduate students from the University of St Andrews were asked to write brief autobiographies, simply detailing important events in their life from childhood to the present. These were combined using a cut-and-paste technique until a single, coherent narrative was obtained. The biographies used in the Experiments a and b were identical except that the biography used in Experiment a stated that the target was accused of committing either an assault or a theft, while that used in Experiment b stated that the target received either a 2i or a 2ii degree. Since the biography used in Experiment a therefore made no mention of criminal behaviour on the part of the target until the final sentence, it could be claimed that this material was somewhat ambiguous, and subjects would be unable to tell whether the target's criminal behaviour was caused by his criminal disposition, or by some special force of circumstance acting at that point of his life. In the latter case, the crime might be thought less stereotypically-linked than in the former case. Nevertheless, it was considered that the advantages of using directly comparable biographies in the two experiments outweighed any advantage that could be

gained through the alteration of the biography used in Experiment a. (The biographies are reproduced in Appendix 11.)

Procedure: One of the six target images was presented to each subject either before or after presentation of the biography. Target images were presented randomly to subjects, with the proviso that all of the images were presented to an equal number of subjects. Each subject was given three minutes to read the biography, and then administered a five-minute, unrelated, filler task, after which s/he was instructed in the use of the Mac-a-Mug programme and asked to regenerate the target image. Once the subject was happy with the regenerated image, the image was printed, and the subject allowed to leave.

The two sets of images thus generated (i.e., in the assault and theft conditions) were submitted to a panel of six independent judges and rated for aggressiveness, honesty, and likeness to the original target.

Results

The raw data by reconstruction condition and stereotype presentation time are presented in Table 4.1. An example of one of the target faces and subsequent regenerations of this face, one for each of the two experimental conditions, are given in Figure 4.1, below.

	Likeness	Aggressiveness	Honesty
Condition			
Assault M	4.492	6.042	4.575
SD	0.922	2.236	3.011
Theft M	4.392	4.150	5.408
SD	2.234	3.149	2.626

TABLE 4.1 - MEANS AND STANDARD DEVIATIONS FOR RATINGS OF REGENERATED IMAGES, "ASSAULT" AND "THEFT" CONDITIONS

All regenerated images were first analysed for their likeness to the target images, to ensure that any effects obtained could not be attributed to the greater likeness of images to target stimuli in one condition than the other. Likeness did not differ significantly with reconstruction condition, $F(1/20) = 0.0323$ $p = 0.8592$, or with time of stereotype presentation, $F(1/20) = 4.3942$ $p = 0.0510$ (see Appendix 7, Table 1).

Neither was there a significant interaction between reconstruction condition and time of stereotype presentation for likeness of regenerated images to targets, $F(1/20) = 0.1516$ $p = 0.7012$ (Appendix 7, Table 1). Hence, any differences in judgments across reconstruction conditions or stereotype presentation times could not be attributed to the greater likeness of regenerated images in one condition than another.

Target

Assault condition

Theft condition

FIGURE 4.1 - TARGET FACE WITH SUBSEQUENT REGENERATIONS, "ASSAULT" AND "THEFT" CONDITIONS

Next, the entire data set was subjected to analysis of variance. A significant main effect of reconstruction condition was discovered, $F(1/20) = 4.6889$ $p = 0.0426$: overall, images regenerated in the "assault" condition were found to be attributed higher scores on the aggressiveness and honesty scales than those regenerated in the "theft" condition. This was not affected by time of stereotype presentation, $F(1/20) = 2.3028$ $p = 0.1448$. A significant interaction was however discovered between reconstruction condition and overall ratings, $F(1/20) = 18.2131$, $p = 0.0004$ (see Appendix 8, Table 1). A Neumann-Keuls test on these data showed that significantly higher ratings of aggressiveness were given to faces regenerated in the "assault" condition than to faces regenerated in the "theft" condition ($Q = 1.892$ $p < 0.01$), a difference of 1.892 on a nine-point scale. Within the "assault" condition, significantly higher ratings were given on the aggressiveness construct than on the honesty construct ($Q = 1.467$ $p < 0.01$). Within the "theft" condition, significantly higher ratings were given on the honesty construct than on the aggressiveness construct ($Q = 1.258$ $p < 0.01$). Faces regenerated in the "theft" condition were rated significantly more honest than faces regenerated in the "assault" condition ($Q = 0.833$ $p < 0.05$), a difference of 0.833 on a nine-point scale.

Time of stereotype presentation was found to have no effect on ratings given, $F(1/20) = 1.5492$ $p = 0.2276$; neither was there a significant interaction between time of stereotype presentation and trait, $F(1/20) = 2.4938)$ $p = 0.1300$. Regenerated images did not receive significantly higher ratings on either trait than the other, $F(1/20) = 0.1065$ $p = 0.7476$ (see Appendix 8, Table 1).

Discussion

It was shown in this experiment that a single piece of stereotypic information embedded in a biography can have a significant effect on the way in which a target face is regenerated, and judgments made by independent judges about these regenerations. When subjects were told that a target was accused of a crime stereotypically linked with aggressiveness, regenerated faces were judged more stereotypically "aggressive" than when they were told that the target was accused of a crime not stereotypically linked with aggressiveness. This finding is in accordance with the hypotheses given above.

The effect was not found to generalise to the condition in which subjects were told that the target was accused of a crime stereotypically linked with honesty. Indeed, judgments of honesty were significantly higher in the "theft" than in the "assault" condition. This point deserves some discussion. It is possible that Macrae and Shepherd's finding of a stereotypic link between theft and

dishonesty does not apply to the present population of judges, or that the physical manifestations of the stereotyped attributes in the regenerated images was too subtle to be used by judges in the experiments reported here. The second point is given some credibility when the actual mean differences in rated honesty between conditions is taken into account. While there was a mean rated difference of 1.892 between ratings of aggressiveness across the two conditions, the difference in rated honesty was just 0.833. On a nine-point scale, this may not represent a very meaningful difference. In this regard, it may be significant to compare the positions of the two pairs of means on a nine-point scale. For the aggressiveness construct, the mean rating fell more than 1.5 points above the mid-point of the scale in the "assault" condition, while in the "theft" condition the mean rating fell 0.35 points below the mid-point (see Table 4.1). Hence, faces regenerated in the "assault" condition were considered (far) more aggressive than the "average" face, while those regenerated in the "theft" condition were considered less aggressive than the average face. Not only was this result statistically significant, therefore, this disparity implies that it is meaningful. This pattern of results does not hold for the honesty construct, in which both means lay above the mid-point of a nine-point scale, implying that, in both "assault" and "theft" conditions, regenerated faces are judged more honest than average. By this argument, therefore, the

difference appears to be less meaningful than the difference in rated aggressiveness across conditions.

An intuitively appealing explanation of the difference in honesty found across these two conditions might be in terms of implicit personality theory (Schneider, 1973) whereby positive traits (e.g., beauty) are stereotypically linked with other positive traits (e.g., honesty), and vice versa, in terms of a global representation of the person as e.g., a "good" or "bad" person. Assault may be viewed by subjects as a more "negative" crime than theft; hence a person accused of assault may be more likely to fall victim to the negative implications of implicit personality theory than one accused of theft. Stereotyping of the target may, therefore, extend into the "honesty" trait: subjects consider a target accused of a dramatic crime like assault such a "bad" character as to be even *less* honest than one accused of theft; and to regenerate his face along those lines. Hence the effect of the "theft" condition in the present experiment may possibly have been swamped by that of the "assault" condition.

This interpretation appears flawed however inasmuch as the mean rating of honesty in the "theft" condition was nevertheless higher than the midpoint of the scale (see Table 4.1). An interpretation of the results of this experiment in terms of implicit personality theory cannot account for why faces regenerated under the "theft" condition should be judged more honest than average. Hence, taking all these factors into

account, it appears likely that an interpretation in terms of the lack of meaningfulness of these data (as proposed above) is most valid.

Two significant results of this experiment have no implications for the hypotheses given above. These were that higher ratings of aggressiveness than honesty were given in the "assault" condition, and that higher ratings of honesty than aggressiveness were given in the "theft" condition. The hypotheses of the present experiment applied only to differences likely to be found between reconstruction conditions, and have no implications for differences found within conditions. Nevertheless, these findings are interesting in themselves, and may have implications for the social stereotype of the two types of criminal considered here.

Again, the effects discovered in this experiment were not affected by time of stereotype presentation. It was hypothesized that, while the effects would hold whether the stereotype was presented before or after biographical information, they would be more pronounced when presented before. This was not found to be the case.

Experiment 2b

In this experiment, it is hypothesized that, where subjects are told that a target received a 2i degree, generated faces should be judged more attractive, and also more intelligent, than when subjects are told that the target received a 2ii degree.

Methodology

The methodology of this experiment was identical with that of Experiment 3a; except that this time the targets were not said to have been accused of any crime, but the class of degree they were said to have received did vary. In one condition, targets were said to have received a 2i degree, and in the other they were said to have received a 2ii degree.

Results

The raw data by reconstruction condition and stereotype presentation time are presented in Table 4.2. An example of one of the target faces and subsequent regenerations of this face, one for each of the two experimental conditions, are given in Figure 4.2.

		Likeness	Intelligence	Attractiveness
Condition				
2i	M	4.083	4.958	4.525
	SD	2.699	2.421	1.494
2ii	M	4.167	4.983	3.442
	SD	1.967	2.812	2.709

TABLE 4.2 - MEANS AND STANDARD DEVIATIONS FOR RATINGS OF REGENERATED IMAGES, "2i" AND "2ii" CONDITIONS

All regenerated images were first analysed for likeness to target images. Neither experimental condition was found to give results more like the

original than the other: $F(1/20) = 0.0167$ $p = 0.8983$. Neither presentation-time affected likeness: $F(1/20) = 0.0429$ $p = 0.8381$. The interaction was also non-significant: $F(1/20) = 0.9671$ $p = 0.3371$ (see Appendix 7, Table 2).

The entire data set was next subjected to analysis of variance. Only one significant main effect was found: ratings on the "intelligence" construct were found to be significantly higher than ratings on the "attractiveness" construct: $F(1/20) = 12.2261$ $p = 0.0023$ (Appendix 8, Table 2).

There was no significant interaction between trait and stereotype condition, $F(1/20) = 3.8503$ $p = 0.0638$, although there was a trend in the expected direction: mean attractiveness rating in the "2i" condition being 4.525, against 3.442 in the "2ii" condition. The traits attributed to the regenerated images were not found to be affected by time of stereotype presentation, $F(1/20) = 0.2092$ $p = 0.6523$. For the main effect of stereotype condition, $F(1/20) = 2.7045$ $p = 0.1157$, and for the main effect of time of stereotype presentation, $F(1/20) = 0.0605$ $p = 0.8082$. For the interaction between these variables, $F(1/20) = 0.4710$ $p = 0.5004$.

Hence, the traits attributed to regenerated images were not affected by the particular stereotype presented, or by time of stereotype presentation; neither did the two stereotype labels differ in their effects when presented at different times.

FIGURE 4.2 - TARGET FACE WITH SUBSEQUENT REGENERATIONS, "2I" AND "2II" CONDITIONS

Together with Experiment 3, the present experiment constitutes one of the first in the literature to test the existance of parallels between facial and whole-body stereotypes: hence the extent of any possible parallel was not known. Therefore it was decided that there would be justification for investigating the effects of the stereotype conditions on trait judgments in separate, smaller analyses of variance in an attempt to isolate experimental effects from the considerable noise likely to be present in the data. Since, once again, no main effect of stereotype presentation time was found, (this reinforced by an independent analysis of variance, see Appendix 9, Table 2) these data were combined for the purpose of this analysis. A significant effect of rated attractiveness was discovered, $F(1/22) = 6.1107$ $p = 0.0216$ (Appendix 10, Table 1): faces generated in the "2i" condition were rated significantly more attractive than those generated in the "2ii" condition. No effect was found for rated intelligence, $F(1/22) = 0.0041$ $p = 0.9497$ (Appendix 10, Table 2).

Therefore, a significant effect of reconstruction condition was discovered in this experiment: faces generated in the "2i" condition were judged more attractive, although not more intelligent, than faces generated in the "2ii" condition. This effect appears to be independent of time of stereotype presentation.

Discussion

The results reported here reinforce the conclusion reached by Shepherd (1991) that stereotypic information can cause bias in the regeneration of facial stimuli. As hypothesized, when subjects were informed that the target face represented an individual who received a 2i degree, the regenerated face was judged more attractive than when they were told that it represented an individual who received a 2ii degree. The strength of this effect was not found to vary with time of stereotype presentation. Results did not appear to be mediated by the rated intelligence of the regenerated images. The fact that the two groups of regenerated images differed in terms of attractiveness but not intelligence is a surprising one, and difficult to explain. Conceivably, it represents the operation of a ceiling effect. The fact that the target was said to have taken a university degree may have been sufficient to establish him as an intelligent character; the relatively small difference in intelligence represented by the award of different degree classes may have been insufficient to cause subjects to regenerate the target stimuli in significantly different ways. The fact that the target biography included no implications about the attractiveness of the target individual, however, may have meant that subjects felt less constrained in terms of the attractiveness of the images which they regenerated.

This argument is supported to some degree by the finding that ratings of "intelligence" were found to be significantly higher across conditions than ratings of "attractiveness". This finding reinforces the contention, made in the Literature Review above, that the constructs, while perhaps highly correlated, should not be considered in any sense interchangeable in social cognitive research. The result may reflect the operation of response-bias: while subjects seem to have been fairly happy to attribute "intelligence" ratings to the target images at above the mid-point of a nine-point scale, (i.e., 4.5,) possibly they may have been less happy to commit themselves to finding one face more attractive than the average (i.e., rating it at above 4.5 on a nine-point scale).

General Discussion

To recap the results of the experiments reported here: Subjects were presented with target faces and told that the targets had been accused of committing either an assault or a theft, or that they had received either a 2i or a 2ii degree. Regenerations made under the "assault" condition were judged more aggressive than those made under the "theft" condition. Regenerations made under the "theft" condition were judged more honest than those made under the "assault" condition: it was argued in the Discussion section of Experiment a that this might be due to the action of implicit personality theory in this case, although a more convincing explantation might be cast in terms of

the lack of meaningfulness of the different ratings in the two experimental conditions.

Subjects told that the targets received a 2i degree regenerated images which were judged more attractive than those regenerated by subjects told that the target received a 2ii degree. This effect was not mediated by the rated intelligence of the regenerated faces: a surprising result which may be attributable to a ceiling effect on the rated intelligence of the targets. In neither Experiment a nor Experiment b did time of stereotype presentation have any effect upon results obtained.

Therefore the hypotheses laid out above were for the most part fulfilled, demonstrating that stereotypic information can affect memory for and judgment of facial stimuli. The potential importance of such an influence in the eyewitnessing situation is quite apparent. As Shepherd *et al* (1987) point out, systematic distortions in memory for facial stimuli have practical consequences for the police when collecting eyewitness descriptions of offenders. Experiment 2a helps highlight the kind of distortion that may arise: witnesses to a crime which is stereotypically linked to aggression may possibly give descriptions of the offender which are biassed towards the stereotypically aggressive. The fact that time of stereotype presentation appears to be immaterial to the distortion produced is of particular importance: implying that even stereotypic information dropped late in the eyewitnessing process - during the police

interview, say - might have the effect of distorting person descriptions. This possibility is discussed in more detail in Chapter 7. The fact that faces regenerated in the "theft" condition were judged more honest than those regenerated in the "assault" condition runs counter to hypothesis, however.

A secondary aim of the present experiment was to provide a standard of comparison for a study of the regeneration of whole-body stimuli. It is conceivable that parallel effects may be found for the stereotyping of body stimuli to those for facial stimuli. For example, the presentation of biographical material which contains the information that a target was accused of the aggressive crime of assault may cause subjects to regenerate a "more aggressive" body-shape. Such possibilities are considered in the following chapter.

CHAPTER 5 - EVERY PICTURE TELLS A STEREOTYPE: THE EFFECT OF STEREOTYPIC INFORMATION ON THE REGENERATION AND JUDGMENT OF WHOLE-BODY STIMULI

"He was really very short and fat
While in the story so skinny and bony."
Billie Holiday: Yankee-Doodle Never Went to Town

Introduction

The present thesis has so far considered the effect of a stereotypic physical feature (body-size) upon the judgment of an ambiguous action, and the effect of stereotypic information upon the regeneration of facial target stimuli. The present chapter completes the first section of this thesis by reviewing the psychological literature on memory for whole-body features, and reporting an experiment which applies the paradigm utilised in Experiment 2 to the regeneration of whole-body stimuli. It is hoped in this chapter to establish the potential importance of whole-body features in the eyewitness situation, briefly to review studies of the relationship between body-type and criminality, and, in Experiment 3, to establish to what extent stereotypic information might affect the regeneration of whole-body stimuli.

Literature Review

The Importance of Whole-Body Information in the Eyewitness Situation: It may be helpful to clarify the reasoning behind studying subjects' ability to interpret or remember whole-body information. As mentioned in Chapter 4, the majority of work in the person-identification literature concerns facial information: psychologists now know a great deal about

witnesses' ability to remember facial information, and the factors which impact upon this ability. There is however reason to believe that other sources of information may also be important. A person description given in everyday life, for instance, is likely to begin with whole-body information: "He's about five foot ten inches tall, *and*...", "She's very thin, *and*...". This habit is not restricted to conversation: it is also present in the works of descriptive writers, of which the following are more or less randomly-chosen examples:

> "Mr Samuel Wilkins was a carpenter, a journeyman of small dimensions, decidedly below the middle size - bordering, perhaps, on the dwarfish. His face was round and shining, and his hair carefully twisted into the outer corner of each eye, till it formed a variety of that description usually known as 'aggerawators'."
> (Charles Dickens, *Sketches by Boz*, p 269)

> "[Captain Black] was a tall, narrow, disconsolate man who moved with a crabby listlessness. He shaved his pinched, pale face every third or fourth day, and most of the time he appeared to be growing a reddish-gold moustache over his skinny upper lip."
> (Joseph Heller, *Catch-22*, p 146)

A few minutes spent studying the fiction shelves in any town library should be sufficient to show that these are not unrepresentative examples of person-description in literature.

The tendency to put body information first suggests that it may be more than just a *part* of person identification: it may conceivably provide a *frame of reference* for the remainder of the description. After all, in trying to communicate something as complex as

a person-description, it makes sense to provide the most basic information first. Body-information may form an infrastructure around which it is possible to build the remainder of a description. Perhaps a phrase of the sort, "He's about five foot ten inches tall" provides the most basic piece of information one needs to know in order to complete a mental representation of a person.

Additionally, in many situations, judgments about other people appear to rest not only on *facial* but also on whole-body information. For example, we are routinely able to recognise friends and acquaintances from a distance or from behind, when facial information is presumably either lost or overshadowed by other sources of information.

It does not take much imagination to see that the ability to recognise individuals on the basis of body-information is of potentially great importance in the eyewitness memory situation (MacLeod, Frowley and Shepherd, in press). This is especially likely when disguise is used: since, although body-information can be hidden (by, e.g., a large coat), it cannot be obliterated in the way that facial information can be obliterated by a simple balaclava.

Body information can be considered a member of the class of context cues investigated in a series of experiments by Cutler, Penrod and their associates, (Cutler, Penrod, O'Rourke and Martens, 1986; Cutler, Penrod and Martens, 1986; Cutler Penrod and Martens, 1987; Cutler and Penrod, 1988). These experiments

consider the situations under which such context cues are most important. Probably the most significant for present purposes is the study reported by Cutler, Penrod and Martens (1986). This paper explicitly studies the effects of "physical characteristic context cues" - a term including, e.g., gait, posture, and voice, as well as physique. One of two types of "array" (conceptually the same as a lineup) was used: subjects received simple slides of suspects' head and shoulders in front and full profile, presented alone or with additional context cues - including voice information and videotapes of suspects entering and leaving a room. They found that "these additional context cues significantly improved overall identification accuracy...when retention interval was long and when subjects were not shown mug-shots after viewing the videotaped crime" (Cutler *et al*, 1986, p 115). And, "[P]hysical characteristic context cues significantly improved identification accuracy when subjects were shown high-similarity line-ups; that is, when the arrays contained several foils who resembled the robber in physical appearance...physical characteristic context cues were most effective when memory for the target was degraded due to encoding factors such as disguise of the robber and weapon focus" (ibid.). This last finding - that reliance on cues other than face information is heavier when face information is made less salient at encoding - may indeed be unsurprising. However it does serve to point up the necessity for psychologists to take

"physical characteristic context cues" into account in the design of experiments.

The importance of contextual variables has also been demonstrated by Thomson and his colleagues (Thomson, 1981; Thomson, Robertson and Vogt, 1982), who exposed subjects to a series of slides depicting target people wearing certain clothing and carrying out certain activities in certain settings. Later, subjects had to pick the targets from a second series of slides. In one condition, the three variables were kept constant. In another, all were changed. Two groups of subjects were used, each being given different instructions on how to treat the memory task. One group was told to maximize the hit rate, i.e., number of *correct* identifications made. The other was told to minimize false alarms, i.e., *in*correct identifications made.

Thomson (1981) found that keeping the variables constant had a massive effect, boosting recognition accuracy from 25 to 89%. It was also found to increase the frequency of false alarms. Thomson argues that "The context elements, background, clothing and actions, are an integral part of the memory information stored by a person when he perceives an event or object" (p 52). An interesting addition to this work is the citation of an unpublished paper by Thomson and Robertson (1980), in which it was found that, for some of the test items, when a new person was shown wearing the same clothing, performing the same action in the same location as

someone in the previous series, subjects seemed to be able to retrieve an "image" of the original event, and confidently reject the new person as having been seen previously. In this finding the positive effects of contextual cues on person memory as discussed by Cutler, Penrod and their associates can clearly be seen. Body information might also be expected to act in this way.

Thomson *et al* (1982) found negative effects on recognition performance when the three types of contextual information were altered; but these effects were not attributable to an *independent* change in any one type of information (Experiment 1 and 2); but rather to a concurrent change in all three (Experiment 3). Unfortunately, this finding does not distinguish between two possible interpretations of the importance of contextual cues in person identification. The effects of changed context may be invisible until a certain threshold has been reached: and in Thomson's particular study this may not have occurred until all three contextual variables had been altered. Alternatively, it may be necessary to alter *all* contextual variables that are present in a given situation before any effect can be seen. Of the two hypotheses, the former appears intuitively more likely; the second cannot however be discounted.

Profound effects on subjects' memory performance or interpretation of information have been found in a number of situations and for a number of different types of contextual information. This has been shown

by e.g., Ekman, Friesen, O'Sullivan and Sherer (1980) for non-facial and vocal cues; by Kozlowski and Cutting (1977) for movement of the body in identifying gender; and by Lindsay, Wallbridge and Drennan (1987) for clothing. All of this work gives good grounds for expecting whole-body information to be of great importance to eyewitness memory.

Despite such demonstrations of its potential importance, however, there is a great lack of research in this area. The majority of the psychological literature associated with whole-body memory and judgment concerns either discussion of Sheldon's system of body classification (this literature is reviewed in Chapter 3) or the body-size perceptions of patients hospitalised with eating disorders (see, e.g., Collins and Plahn, 1988). A small literature also exists on the relationship between physique and criminality. This literature is discussed briefly in the following sub-section.

Whole-Body Features and Their Relationship with Criminality: As discussed in Chapter 2, there is little support in the psychological literature for Sheldon's theory of constitutional psychology: it appears unlikely that psychological variables have a direct relationship to the physical variables of body-size and shape. Nevertheless, some researchers have found evidence for a relationship between body-shape and criminality. Glueck and Glueck (1955) conducted a careful study in which delinquent boys were matched

for age, intelligence, racial origin, and residential background with non-delinquent boys. The delinquent group was found to be considerably more mesomorphic and less ectomorphic than the non-delinquent group. Epps and Parnell (1952) conducted a study of the body-shapes of young women in Borstal or university, finding those in Borstal to be heavier, more muscular, and fatter. Eysenck (1964), in gathering support for his theory of the relationship between somatotype and criminality, concludes on the basis of such data that criminals have a distinct tendency towards extraversion; a trait which he believes to be related to the mesomorphic somatotype. And Gibbens (1963) carried out an extensive study of the body-builds of "juvenile delinquents" detained in Borstal institutions, revealing that criminals tended to fit the category of "endomorphic mesomorphs": being athletic in build, i.e., more "stocky and muscular" than "fat". They also showed a tendency towards extraversion and the related trait of impulsiveness.

Interesting as these conclusions may be, however, the amount of research which has been conducted in the area is limited, and it would be dangerous to draw any hard-and-fast conclusions from it. Also, these studies appear to leave open the question of direction of causation. Regarding Glueck and Glueck's (1955) study, for instance, it may be that the two groups of subjects - the "delinquent" and "non-delinquent" boys - did not in fact differ on any psychological trait. Conceivably, it is easier to become "delinquent" if

one has a mesomorphic - more muscular and athletic - body-type. Alternatively, it is equally conceivable that the boys with the more mesomorphic body-type were simply trying to fulfill a social stereotype of mesomorphs when they became "delinquent". This hypothesis is lent extra weight by studies which illustrate the stereotypical relationship between certain body-types and psychological characteristics, reviewed in Chapter 3. It is such a relationship that Experiment 3 aims to investigate.

Stereotypes and Whole-Body Information: Of particular importance to the present study is the effect of stereotypes on perception of and memory for whole-body information. It is conceivable that people who possess certain body-types may be subject to biasses similar in kind (if less extreme) to those experienced by, e.g., members of minority ethnic groups. The studies which address this issue are reviewed in Chapter 3.

Judging from the powerful effects demonstrated for stereotypes in the literature, (from studies of biographical material by Snyder and Uranowitz, 1978c, to facial information by Macrae and Shepherd, 1989a, to actions performed in video-recordings by Duncan, 1976,) one would be naïve to expect stereotypic information to have no effect in this instance. However, this particular area of stereotyping research has received no attention in the literature, despite the fact that it could have powerful implications for

eyewitness' memory of physiques. Indeed, the social cognition literature might lead one to predict almost any effect one might care to mention for whole-body memory. Classically, there are two strands to this research: investigators have considered the effects of stereotypes presented at encoding of to-be-remembered information and (less frequently) those of stereotypes presented at retrieval. This literature will be exhaustively reviewed later in the present thesis, but for the moment a more or less random sampling might give an idea of its contradictory nature. Rothbart (1981) found that stereotype-consistent information was better recalled than inconsistent or irrelevant information; O'Sullivan and Durso (1984) found that consistent information was best remembered, but only when *very* inconsistent information was also presented; Hastie and Kumar (1979) found superior recall for inconsistent information; and a similar result was found by Crocker, Hannah and Weber (1983).

The social cognition literature gives no *a priori* grounds to predict what type of information - consistent or inconsistent with the physical stereotype of the criminal - will be better remembered by an eyewitness; or indeed whether *either* will be better remembered. Given that there is potential here for elucidation of extra information from an eyewitness, as well as potential for an eyewitness' account of a crime to undergo systematic distortion, it seems important that researchers with interests in

the area should begin systematic investigation of the area.

A smaller literature exists on the effects of stereotypes presented at retrieval: one which centres around the ongoing controversy concerning Snyder and Uranowitz' seminal paper on the effects of a "lesbian stereotype" on memory for a biography of a fictional character, Betty K. Calls for greater emphasis in the literature on stereotypes presented at retrieval (see, e.g., Bodenhausen and Lichtenstein, 1987; Bodenhausen and Wyer, 1985) have unfortunately gone all but unanswered; because of either the thorny nature of the area, or the fact that the majority of models of effects of stereotypes on memory take little account of the possibility of this effect (see Chapter 6). Bodenhausen and Lichtenstein write:

> "...there is a great deal of research documenting the importance of encoding operations and organizational processes, but there is a relative dearth of evidence provided by social cognition research documenting retrieval effects" (p 879).

The present study will take the opportunity partially to fill this hole in the literature by considering the effects of stereotypes presented before and after target information.

A Paradigm for Research into the Whole-Body Arena:

The literature on the processing of target faces was reviewed in Chapter 4. This processing appears to be affected by such categorical information as age, attractiveness, gender, intelligence, occupation,

Brigham, 1971; Cash, 1981; Keating, 1985; Klatzky, Martin and Kane, 1982 a and b; MacArthur, 1982; MacArthur and Apatow, 1983). Parallels can be drawn between this work and that on body-size stereotyping by Christiaansen, Sweeney and Ochalek (1983); Ertel and Prodhöhl (1969); Jansen and Horowski (1980); Lerner and Moore (1974); and Wilson (1968), discussed in Chapter 3. Indeed, it is possible to draw an analogy from the face-processing literature to the whole-body literature. "Upper face features have...been shown to be more important in face recognition (see Fisher and Cox, 1975; Garneau, 1973) and one might suppose that upper body features may be more important to body recognition than lower body features" (MacLeod *et al*, in press). Therefore it is conceivable that the effects of stereotypic information on the regeneration of facial stimuli may also hold when whole-body stimuli are used. Experiment 3 is a conceptual replication of Experiment 2, this time using whole-body information rather than facial information as target stimuli.

An Outline of Experiment 3

The effects demonstrated in Experiment 2 provide a good starting point for the investigation of the effect of stereotypes on whole-body memory. It was important to enter this area with the intention of investigating more or less robust phenomena. Were a wholly new experiment to be designed, then one could not know whether to explain the results in terms of

the action of stereotypes on whole-body memory in general, or that one instance in particular. However, results (whether positive or negative), in a paradigm with strong connections to those used previously, would be far more easily explained.

Particularly germaine here is the fact that Experiment 1, while demonstrating a powerful effect of stereotypic information on judgment of an action, utilised a paradigm in many ways the reverse of that used here. The present experiment considers the effect of stereotypic information upon the reconstruction of whole-body stimuli, in an effort to demonstrate the generalisability of whole-body stereotypes to other situations of interest in the eyewitness situation. Hence, an experiment was designed with the intention of seeing how far the results of Experiment 2, and those of Macrae and Shepherd (1989a) and Shepherd (1991), held for whole-body information.

Especially important in the eyewitness situation is the effect of stereotyping with respect to criminal occupation. The applicability of criminal stereotypes to facial stimuli has been demonstrated in Experiment 2, and by Macrae and Shepherd. This forms the main area of interest in the present experiment; however a second experiment is also reported in which subjects are given information about the degree class awarded the target individual, as in Experiment 2b. The purpose of this manipulation is twofold: to check the validity of generalising from facial to body stimuli

by utilising conditions which have already been shown to have ramifications for facial stimuli; and to extend the presently limited knowledge of the effects of stereotypic information on body stimuli.

The present experiment considers not only the effects of stereotypes on reconstruction of body stimuli, but also the effects of time of presentation of stereotypic information, as in Experiment 2.

Since the two conditions in this experiment are not directly related, they are reported separately as Experiments a and b.

Experiment 3a

The hypotheses for the present experiment are presented below:

1. Where subjects are told that the target individual committed an assault, then the body-shape regenerated should be judged more "aggressive" than when subjects are told that the target committed a theft. Where subjects are told that the target committed a theft, then the body-shape regenerated should be judged less "honest" than when subjects are told that the target committed an assault. Extrapolating from the results of Experiment 1, which helped to establish that a larger body-size may stereotypically be associated with a more "aggressive" personality, it appears likely that a person involved in an assault will stereotypically be thought of as having a larger body-size than a person involved in a

theft. However, it is less immediately obvious what body-type may stereotypically be associated with the "honest/dishonest" dimension. Nevertheless, the "theft" condition is included in the present study for a number of reasons. First, it acts as a control condition for the "assault" manipulation. Second, it allows comparison to be drawn with Experiment 2a, in which the "theft" manipulation was utilised, and hence allows us to increase our knowledge of the similarities and dissimilarities between facial and whole-body stereotyping. Finally, the manipulation helps to establish whether or not there does in fact exist a social stereotype of an "honest/dishonest" body-type.

2. Where subjects are told that a target committed an assault, then a larger body-shape should be generated than when they are told that the target committed a theft.

3. These effects should be less pronounced, although still present, when stereotypes are presented *after* to-be-remembered information than when they are presented *before* to-be-remembered information.

Methodology

Subjects: Subjects in this experiment were forty-eight undergraduate and postgraduate volunteers, who took part for a small fee. Subjects were aged between 18 and 27 years. Thirty-one subjects were female, seventeen were male.

Apparatus: For regeneration of target photographs, a Kowa 2x anamorphic lens was used. This lens attaches to the front of a slide-projector and, when turned through 360°, produces consonant size distortions in the image produced. The lens has been used extensively in research into the self-body images of anorexic patients (see, e.g., Collins and Plahn, 1986,) but appears never before to have been applied to a more strictly social-cognitive setting.

Materials: Using a Pentax Asahi SP1000 loaded with Kodachrome Gold 400 film, photographs were taken of eight target persons. Three photographs were taken of each target: one with the camera held on a tripod perpendicular to the floor, one with the camera tilted 15° to the left, and one with the camera tilted 15° to the right. This was done so that during the reproduction phase subjects would not be able to use the angle of the image as an indication of the accuracy of their work: the anamorphic lens altering the angle of a projection as well as its apparent size. All targets were males aged between 24 and 27. All photographs had heads and facial features occluded and were taken against a plain wall.

Next, all body-shapes were presented to five independent judges (Honours Psychology undergraduates from the University of St Andrews) who rated them for intelligence, attractiveness, honesty, and aggressiveness. Only targets balanced on these dimensions were used as target stimuli in the

experiment. Four targets fulfilled this criterion (see Appendix 12, Table 1).

A two-hundred word biography was constructed in two stages: four randomly-selected postgraduate students from the University of St Andrews were asked to write brief autobiographies, simply detailing important events in their life from childhood to the present. These were combined using a cut-and-paste technique until a single, coherent narrative was obtained. Two versions of the biography were used in the present experiment. In one version, the subject of the biography was said to have been accused of committing an assault; in the other of committing a theft. (These biographies are reproduced in Appendix 11).

Procedure: A target image was presented to each subject, via Kodak Carousel slide-projector, either before or after presentation of the biography purportedly relating to the target. The target images were presented randomly to the subjects, with the proviso that each target image was seen by two subjects, one in the "assault" condition, and one in the "theft" condition. The subject was then given a brief (five minute) unrelated filler task, after which s/he was asked to reproduce as best s/he could the body-image seen previously. A photograph was then taken of this regenerated image, using a camera mounted on a tripod at a fixed distance of 2.5 metres from the screen upon which the image was projected. The two sets of photographs thus generated (i.e., in

the assault or the theft condition) were then submitted to a panel of independent judges and compared for ratings of aggressiveness, honesty, body-size, and likeness to the original images. Measurements were also taken of the *actual* mean percentage distortion produced by subjects in each condition. This was performed to test whether judges' interpretations of differences between the body-shapes regenerated across conditions could be attributed to the existence of actual differences across conditions.

Results

a) Actual Distortions of Regenerated Images: The actual percentage distortions by reconstruction condition are presented in Table 5.1, below.

Condition	Assault	Theft
Mean % Distortion	+25.887	+30.148
SD	11.430	1.617

TABLE 5.1 - MEAN PERCENTAGE DISTORTION OF REGENERATED BODY-STIMULI x RECONSTRUCTION CONDITION, EXPERIMENT 3a

Immediately noticeable from the data presented above is the difference between the two standard deviations. The fact that the standard deviation was far larger in the assault than in the theft condition is probably not best explained in terms of outliers in the assault condition. Table 5.3, reproduced under Experiment b, below, gives the means and standard deviations which were generated under 2i/2ii degree class conditions. Under each condition in Experiment b, the standard

deviation is comparable to that found in the assault condition here. Therefore, the standard deviation in the assault condition does not appear to be the unusual case. The difference here should therefore be explained by reference to the greater clustering of data points about the mean in the theft condition. There appears to have been far greater agreement among subjects about the appearance of a thief's body than about the body of someone accused of assault.

Analysis of variance of these data showed no significant effect of stereotype reconstruction condition on percentage distortion of regenerated images, $F(1/47) = 3.1934$ $p = 0.804$ (see Appendix 12, Table 2). The trend of the data is in fact in the opposite direction from that predicted: it was expected that the crime stereotypically linked with body-size (assault) would lead to the production of larger images, not vice versa. Any significant results obtained in this experiment could not be attributed to actual differences in reconstructed body-size across the two conditions.

b) Ratings of Regenerated Images: The raw data by reconstruction condition and stereotype presentation time are presented in Table 5.2.

All regenerated items were analysed for likeness to the original images, to check whether any differences found for the other variables could be explained in terms of greater likeness to the original stimuli in one condition than the other.

		Likeness	Aggressiveness	Honesty	Size
Condition					
Assault	M	4.642	5.333	5.250	5.692
	SD	1.315	1.055	0.627	1.025
Theft	M	5.083	4.900	5.159	4.725
	SD	0.871	1.421	0.570	0.835

TABLE 5.2 MEAN RATINGS AND STANDARD DEVIATIONS OF REGENERATED IMAGES x STEREOTYPE PRESENTED x STEREOTYPE PRESENTATION TIME, EXPERIMENT 3a

Likeness was found to differ significantly with stereotype condition: $F(1/44) = 8.9213$ $p = 0.0046$, and with the time of presentation of stereotypic information (i.e., before or after presentation of the target image): $F(1/44) = 6.1487$ $p = 0.0171$. The interaction between condition and stereotype presentation time was not significant: $F(1/44) = 1.8294$ $p = 0.1831$ (Appendix 13, Table 1).

Next, a large-scale analysis of variance was conducted on the full data-set to indicate the significance of any main effects or interactions. No significant effect was found. For stereotype condition, $F(1/44) = 2.2813$ $p = 0.1381$. For time of stereotype presentation, $F(1/44) = 0.0063$ $p = 0.9370$. Hence, stereotypes did not have different effects with different presentation times. For the interaction between stereotype condition and presentation time, $F(1/44) = 0.2429$ $p = 0.6245$. Regenerated images were not found to be given significantly higher ratings on any one trait than any other, $F(2/88) = 2.7052$ $p =$

0.0724; neither was this affected by stereotype condition: $F(2/88) = 0.9237$ $p = 0.4008$, or stereotype presentation time: $F(2/88) = 1.0128$ $p = 0.3674$, (Appendix 14, Table 1). Therefore, the stereotype presented appeared to have no significant effect upon traits present in reconstructed body-images, and neither did stereotypic information appear to have different effects when presented at different times.

The present experiment constituted one of the first in the literature to use the whole-body as target stimuli: hence the magnitude of the expected effects was not known. Therefore, as in Experiment 2b, it was decided to investigate the effects of the stereotype conditions on trait judgments in separate, smaller analyses of variance in an attempt to isolate experimental effects from the noise that may be be present in the data. Since time of stereotype presentation had no independent effect upon ratings given, (an effect also supported in a separate analysis of variance, see Appendix 15, Table 1,) these data were combined for analysis.

No significant effect of stereotype condition was found for aggressiveness ratings: $F(1/46) = 2.3581$ $p = 0.1315$, or for honesty ratings: $F(1/46) = 0.0354$ $p = 0.8516$. However, bodies regenerated under the "assault" condition were rated as being significantly *larger* than those regenerated under the "theft" condition, in line with the experimental hypotheses: $F(1/46) = 4.1053$ $p = 0.0486$ (Appendix 16, Tables 1 - 3).

Discussion

This experiment found that a piece of stereotypic information embedded in a two-hundred word biography can significantly affect the way in which a target body-shape is regenerated. When a target was said to have committed a crime stereotypically linked with aggressiveness, regenerated body-shapes were rated by independent judges as significantly larger than when the target was said to have committed a crime linked with dishonesty but not aggressiveness (theft). Given the significant correlation reported with Experiment 1 between perceived body-size and perceived aggressiveness, this is precisely the result expected.

Two results that are somewhat inconsistent with this finding were also achieved however. No actual difference in the size of the regenerated images was discovered. Hence, judges' interpretations of this variable appear to have been based on some factor other than the actual size of the images they were studying. It is tempting to argue that this factor might have been the perceived aggressiveness of the images: but if this is so, then it is perhaps odd that no significant effect was discovered for the aggressiveness trait *itself*. The correlation found between body-size and aggressiveness is admittedly quite small, even though it is highly significant ($r = 0.1359$ $p < 0.001$). Nevertheless, one would have expected the stereotypic information that the target committed an assault to have had a direct effect upon

aggressiveness *per se*, and not solely on a correlate of aggressiveness. Conceivably, this effect may be traceable to an artifact of the experimental materials used. As mentioned in Chapter 4, the biographies presented to the subjects in the present experiment may suggest that the cause of the crime committed by the target was not a dispositional one. Subjects may not therefore have felt that the target was a dispositionally aggressive person, and hence did not generate body-images which reflected stereotypic attributes of a dispositionally aggressive target. It should however be noted that there was a trend in the expected direction for the aggressiveness construct: mean rated aggressiveness in the "assault" condition being 5.425, against 4.875 in the "theft" condition, a difference of 0.55 on a nine-point scale.

It is evident that only the "assault" condition had any effect upon judges' ratings of traits in the regenerated body-shapes. Where targets were said to have been accused of theft, no significant different in rated honesty was found (a mean of 5.025 in the "theft" condition against a mean of 4.983 in the "assault" condition). As mentioned above, it is by no means immediately obvious just what the stereotypically "honest/dishonest" body-shape might consist of; and, indeed, one aim of including this condition in the present experiment was to establish to what extent whole-body stereotypes parallel facial stereotypes, and whether there is indeed any such thing as a social stereotype of an "honest/dishonest"

body-type. The results of this experiment give cause to doubt whether any such stereotype does in fact exist. This is unsurprising when it is considered that the "theft" manipulation in Experiment 2a did not have a significant effect upon the rated honesty of regenerated faces, despite the fact that effects might be thought more likely to obtain when facial than whole-body stimuli are used.

It could, however, be argued that another factor played the most important role in these results. It was found that the mean likeness of regenerated images to original stimuli was higher in the "theft" than in the "assault" condition. Conceivably, the main effect for body-size found here was caused by greater similarity in one condition, (or dissimilarity in the other,) to the original stimuli. Alternatively, a third variable may have had affected both body-size and similarity. It is conceivable that one of the two stereotypes presented in this experiment might have acted as a retrieval schema for the target information (see, e.g., Snyder and Uranowitz, 1978c): however this is a very tenuous conclusion to draw from the limited evidence available here. This possibility is discussed in far greater detail in Experiments 7 - 8, below.

Experiment 3b

This experiment is a conceptual replication of Experiment 2b, utilising body rather than facial information for target stimuli. Two conditions were

run, in which subjects were told that the target individual received either a 2i or a 2ii degree. It is hypothesized that, where subjects are told a target received a 2i degree, then a more attractive body-shape should be generated than when subjects are told the target received a 2ii degree. Since this effect may be confounded by the effect of perceived *intelligence*, (see the discussion of this possibility in Chapter 3,) ratings of intelligence were also taken. It is hypothesized that body-shapes generated under the "2i" condition will be judged more intelligent than those regenerated under the "2ii" condition.

Methodology

The methodology for this experiment was identical with that of Experiment a, above, except that the biography did not mention that the target had been accused of any crime, but instead varied the class of degree said to have been awarded, i.e., either a 2i or a 2ii degree.

Subjects: A further forty-eight subjects were used in this experiment. These were undergraduate and postgraduate student volunteers, who participated for a small fee. Twenty-nine subjects were female; nineteen were male.

Results

a) *Actual Distortions of Regenerated Images:* Mean percentage actual distortion and standard deviations are presented in Table 5.3, below.

Condition	2i	2ii
Mean % Distortion	+26.835	+26.459
SD	10.093	10.766

TABLE 5.3 - MEAN PERCENTAGE DISTORTION OF REGENERATED BODY-STIMULI x RECONSTRUCTION CONDITION

Analysis of variance showed no significant difference in the actual size of regenerated body-image across condition, $F(1/46) = 0.0156$ $p = 0.9011$ (Appendix 12, Table 3). Therefore, any difference found in ratings of these images could not be caused by actual differences in image size.

b) *Ratings of Regenerated Images:* Raw data by reconstruction condition and stereotype presentation time are presented in Table 5.4.

		Likeness	Intelligence	Attractiveness
Condition				
2i	M	4.650	5.133	4.671
	SD	1.344	0.863	1.779
2ii	M	4.867	5.017	4.317
	SD	2.000	0.676	0.698

TABLE 5.4 MEAN RATINGS AND STANDARD DEVIATIONS OF REGENERATED IMAGES x STEREOTYPE PRESENTED x STEREOTYPE PRESENTATION TIME, EXPERIMENT 3b

Again, all regenerated images were analysed for likeness to the original images presented. Neither stereotype condition was found to affect the likeness of the images regenerated: $F(1/44) = 0.3902$ $p = 0.5354$. Neither stereotype presentation time (before or after presentation of target material) affected likeness: $F(1/44) = 1.3923$ $p = 0.2444$. The interaction was also non-significant: $F(1/44) = 0.5829$ $p = 0.4492$ (see Appendix 13, Table 2).

Next, a single large analysis of variance was performed on the entire data set - i.e., the ratings obtained for both dependant variables, in both stereotype conditions, and at both presentation times, (i.e., either before or after presentation of the target stimuli,) - to indicate whether any main effects or interactions appeared to be significant. This was not found to be the case. For both the main effects of stereotype condition and time of stereotype presentation, $F(1/44) = 0.3199$ $p = 0.5746$. For the interaction between stereotype condition and time of presentation, $F(1/44) = 0.0943$ $p = 0.7603$. Hence, as in Experiment 2b, traits attributed by the independent judges to the regenerated images were equally unaffected by the stereotype presented and the time of stereotype presentation. Neither was it found that the stereotype labels had different effects when presented at different times, as it was expected they would. Only one main effect was found to be significant: ratings on the "intelligence" construct

were significantly higher than ratings on the "attractiveness" construct: $F(1/44) = 19.9882$ $p = 0.0001$. This was not affected by stereotype condition, however: $F(1/44) = 0.0493$ $p = 0.3629$; or by time of stereotype presentation: $F(1/44) = 0.0251$ $p = 0.8748$. (See Appendix 14, Table 2, for analysis of variance tables.)

Again, separate, smaller analyses of variance were also employed. There being no independent effect of time of stereotype presentation, (this backed up by analysis of variance performed on this data exclusively - see Appendix 15, Table 2,) these data were combined for analysis. No main effect of stereotype condition was discovered on the dependent variable of rated intelligence: $F(1/46) = 0.2108$ $p = 0.6483$ (Appendix 16, Table 4), or on rated attractiveness: $F(1/46) = 1.5196$ $p = 0.2239$, (Appendix 16, Table 5).

Therefore no effect of stereotype condition could be isolated on judgments made about regenerated bodies in this experiment.

Discussion

It was hypothesised that, in the present experiment, similar effects would be seen to hold for body stimuli as had previously been discovered for face stimuli by Shepherd (1991), and in Experiment 2b: i.e., that when a target was said to have received a 2i degree, regenerations of that target's physical appearance would be judged more attractive then when the target

was said to have received a 2ii degree. It was also hypothesized that such a result might be mediated at least to some extent by the effect of intelligence across the two conditions. Neither hypothesis was supported. Body-shapes reconstructed in the 2i condition were not judged significantly more attractive or intelligent than those reconstructed in the 2ii condition. Hence, in this respect at least, the reconstruction of body stimuli appears to be less susceptible to stereotypic bias than that of face stimuli.

It should, however, be mentioned that in both conditions the trend of results was in the expected direction. Mean rated attractiveness in the 2i condition was 4.671, against 4.317 in the 2ii condition; mean rated intelligence was 5.133 in the 2i condition, against 5.017 in the 2ii condition. Nevertheless, the fact that these trends were not found to be significant highlights the fact that the effects under investigation here are of a particularly subtle and elusive nature, or may not exist at all. Hence the results of the present experiment indicate a limitation to the possible parallels that may be drawn between facial and whole-body stimuli as far as social stereotyping is concerned.

The fact that ratings for "intelligence" were found to be significantly higher across conditions than those for "attractiveness" also deserves mention, since it appears to indicate that neither judgment is wholly reliant on the other; as mentioned in the

General Discussion section of Experiment 2. While the two judgments may be highly correlated - at least in much research which has used facial target stimuli - the constructs are certainly not interchangeable, and, in the present area at least, are likely to have a correlational rather than causal relationship. Judges appear to have been fairly happy to ascribe "intelligence" ratings of greater than five on a nine-point scale to the regenerated images used here, but loathe to rate them at above about four-and-a-half for "attractiveness". Possibly (as mentioned in the Discussion section of Experiment 2b) this indicates some degree of response-bias in a population of judges who did not want to commit themselves to finding particular body-shapes "more attractive" than others.

General Discussion

Briefly to recap the results of the experiments presented here: subjects were asked to regenerate a target body-image which they had viewed either before or after receiving a piece of stereotypic information concerning the target. In the first experiment, this information pertained to the nature of a crime committed by the target; in the second, to the class of degree achieved. It was found that class of degree did not affect the rated intelligence or attractiveness of the regenerated image, as had been hypothesized from the results of Macrae and Shepherd (1989a), Shepherd (1991), and Experiment 2 in the present thesis. Stereotypic information concerning a

crime committed by the target was found to affect the rated body-size of the regenerated image (a correlate of aggressiveness), although it was not found to affect rated aggressiveness itself. Information concerning a crime stereotypically linked with dishonesty (theft) was not found to affect the rated honesty of the regenerated image. As mentioned in the Methodology section, this might be related to the difficulty of establishing the possible nature of a stereotypically "honest/dishonest" body-shape. The present experiment also showed a significant effect of stereotype condition upon rated likeness of the regenerated image to the target image: this was found to be greater in the "theft" condition than in the "assault" condition. It is possible that the other results achieved in this experiment are explicable in terms of this effect; although the mechanism involved is far from obvious. Conceivably one piece of stereotypic information presented here acted as a retrieval mnemonic: an intriguing and potentially important area for investigation in the eyewitness field, and one which will receive attention later in the present thesis (see Chapters 7 and 8).

A surprising result of the experiments presented here was that in neither case did time of stereotype presentation affect the results obtained. The majority of the psychological literature concerning the effects of stereotypic information on memory would lead one to expect it to have a more powerful effect when presented before the target information than when

presented afterwards. Indeed, many models of the effects of stereotypic information hold that it will be ineffective when presented after target information. As pointed out by Bodenhausen and Lichtenstein (1987) this may reflect a bias in the literature (which rarely utilises this paradigm) more than it does the genuine action of stereotypes. The present experiment goes some way towards supporting this claim. This issue is considered in detail in Chapter 6.

Another aim of this experiment was to discover how far the effect of stereotypic information on body material might parallel that on facial material. In Experiment 3a it was found that, in the "assault" condition, larger body-sizes were regenerated than in the "theft" condition: a trait which had previously been found to be positively correlated with perceived aggressiveness. To a limited extent, this parallels the result achieved in Experiment 2a, in which it was found that regenerations made under the "assault" condition were judged more aggressive than those made under the "theft" condition. However, the body-shapes regenerated under the "assault" condition were not judged more aggressive than those regenerated under the "theft" condition. No effect of degree class was found on regenerated body stimuli in Experiment 3b, although it was found in Experiment 2b that regenerations made under the "2i" condition were judged more attractive than those made under the "2ii" condition. Taking these results together, it cannot

easily be argued that the effects of stereotypes on regeneration of whole-body materials are in any simple way parallel to those on regeneration of facial materials. However, while the effects achieved in Experiment 3 are weaker than those achieved in Experiment 2, the fact that a significant result was achieved in Experiment 3a implies that further work should be carried out in this area, in an attempt to establish the extent to which stereotypes of whole-body features parallel those of facial features, and under what circumstances they will prove important.

The present study provides some evidence that stereotypes can bias interpretation of target body information. However, the evidence is not very strong: this experiment indicates that the effects may be limited to the situation in which a target is accused of committing a crime stereotypically associted with aggressiveness. Experiment 1 indicates that a relationship may exist between such a crime (in this case, assault) and the stereotypic body-shape of the perpetrator. The stereotypic relationship between dishonesty and body-shape may be rather less explicit; indeed the present experiment gives little reason to suppose that such a relationship does in fact exist. Hence it seems likely that an eyewitness' account of the whole-body features of a criminal might be biassed by the stereotypic nature of the crime only in a very limited range of situations.

CHAPTER 6 - STEREOTYPES AND MEMORY: A REVIEW OF THE LITERATURE

Introduction

Many studies in the literature have found stereotypic information to have a distorting effect upon memory of target material. Hence it is important that the present thesis attempt to isolate the implications of stereotypic information for eyewitness memory. Most obvious of these implications are the potential negative effects. Consider the following scenario, derived from the famous study of rumour by Allport and Postman (1945): A man, travelling on the London Underground, witnesses an attack by a black man on a white man, during which blows are exchanged. On the premise that such an incident would come to the attention of the police, suppose that the witness be interrogated by a (white) police officer - himself a social stereotype of the racist white policeman - who during the course of the interview lets slip the stereotypic information that blacks are likely to carry knives with them wherever they go. The witness returns to the police station for a second interview, during the course of which he misremembers the black man as having been armed with a knife. Such distortion of memory may have powerful negative implications for the man who committed the assault.

This is an example of the way in which stereotypic information may influence eyewitness memory for an event: sometimes at great cost to the individual. Certainly the example is quite extreme: but it is not

hard to imagine the potential for such incidents within the criminal justice system, given the number of interviews and interrogations that take place.

There is also, of course, the potential for stereotypic information to have more positive effects: the encoding of an event in terms of memorial schemata (Alba and Hasher, 1983) - of which stereotypes may be considered an example (Hamilton, 1979; Macrae, in press) - may lead to more extensive cognitive processing of certain items, and subsequent ease of retrieval for those items (Bodenhausen, 1988). Also, when stereotypic information is presented at retrieval, there is the possibility of its acting as a retrieval schema for certain aspects of an event (Snyder and Uranowitz, 1978c; Snyder, 1981).

The first part of the present chapter concentrates upon social cognitive models of the effects of stereotypes on memory. Again, this is not to deny the importance of the social psychological perspective, but to recognise the importance of the social cognitive approach - especially at this early stage in research, at which it is helpful to work within the framework of the kinds of models which the approach provides. Two research areas can be identified in the social cognitive literature: stereotypic information can be presented either at encoding or during retrieval of target information. These areas are considered separately.

Stereotypes Presented at Encoding

Models of social memory have for the most part ignored the effects of stereotypic information presented at retrieval; being concerned almost exclusively with its effects when presented at encoding. This in turn has generated a disproportionate amount of research into these effects; and, as the literature has grown, the models have become more complex. Few early studies in this area made explicit claims to proposing actual models of stereotypes and memory. Those which did (e.g., Cohen, 1977/81; Cantor and Mischel, 1977; Graesser, Gordon and Sawyer, 1979; Graesser, Woll, Kowalski and Smith, 1980; Hastie, 1980) tended to propose principles (such as schema-consistency) which were conceptually so simple that, as the literature grew, they became unable to account for the large number of disparate findings that appeared. No model postulating a simple one-to-one relationship between, say, consistency and memorability could possibly account for the range of findings reported in the literature (see Table 6.1).

Additionally, methodological factors should be taken into account. Srull (1984) outlines many methodological considerations for work in the social cognition area, remarking that divergent approaches may yield divergent results: in free recall, for example, analysis of the actual words used by a subject may yield different results from analysis of the *meaning intended*. As early as 1979, Dreben, Fiske

TABLE 6.1 - META-ANALYSIS OF THE EFFECTS OF STEREOTYPES ON MEMORY

EXPERIMENTER	YEAR	Ss	MATERIAL	PARADIGM	B/A	EFFECT ON MEMORY
Bellezza & Bower	1982	UG	Prose	Memory Recall/ Recog.	B	Consistency effect; lower recall for irrelevant info.
Bodenhausen	1988	UG	Prose	Judgment Recall	B	Consistency effect both before and after
Cantor & Mischell	1977	UG	Prose	Memory Recall	B	Consistency effect
Cohen	1977	UG	Video	Memory Recall	B	Consistency effect
Cohen	1981	UG	Video	Memory Recog.	B	Consistency effect
Graesser et al	1979-82	UG	Prose	Memory Recall	B	Consistency effect; lower recall for irrelevant info
Halpern	1985	UG	Prose	Memory Recall	B	Consistency effect
Hamilton & Rose	1980	UG	Prose	Memory Recall	B	Consistency effect
Koblinsky & Cruse	1981	4 - 10 yr olds	Prose	Memory	B	Consistency effect
Martin & Halverson	1983	5 - 6 yr olds	Pictures	Memory Recog.	B	Consistency effect
Rothbart	1981	UG	Prose	IF Recall	B	Consistency effect; equal levels irrelevant/ inconsistent
Stangor&Ruble	1989	UG	Prose	Memory	B	Consistency effect information
Rothbart, Evans & Fulero	1979	UG	Prose	Memory Recall	B/A	Consistency effect; equal recall for inconsistent & irrelevant material
Roberts	1985	meta - analysis		Recall	B	Little variance due to consistency
Crocker, Hannah & Weber	1983	UG	Prose	Memory Recall	B	Inconsistency effect
Hastie & Kumar	1979	UG	Prose	IF Recall	B	Inconsistency effect. Lower recall level for irrelevant than consistent or inconsistent info.
Hashtroudie & Mutter	1980	UG	Prose	IF Recall	B	Replicated Hastie & Kumar
Hemsley & Marmurek	1982	UG	Prose	IF Recall	B	Replicated Hastie & Kumar; but effect depends on relative set size
O'Sullivan & Durso	1984	UG	Aural	Memory Recall	B	Very inconsistent information made consistent info. more memorable. Lower recall for

Srull	1981	UG	Prose	IF Recall/ Recog.	B	Inconsistency effect in recall no effect in recog. Lower recall/recog. for irrelevant info.
Hastie	1980	UG	Prose	Memory Recall	B	Inconsistency effect. Lower recall for irrelevant info.
Wyer & Gordon	1982	UG	Prose	Mem/IF Recall	B	IF-inconsistency effect Memory-little effect either way
Wyer, Bodenhausen & Srull	1984	UG	Prose	IF Recall/ Recog.	B/A	Inconsistency for recall when presented before Consistency for recall when presented after
Bower Black & Turner	1979	UG	Prose	Memory Recall	B	Lower recall for irrelevant info.
Snyder & Uranowitz	1978	UG	Prose	IF Recog.	A	Consistency effect
Clark & Woll	1981	UG	Prose	IF Recog.	B/A	No effect of stereotype in either condition
Bellezza & Bower	1981	UG	Prose	IF Recog.	A	What little effect there is, is due to response bias
Lutz	1983	UG	Prose	IF Recall/ Recog.	A	Consistency effect

Key - "Paradigm" column gives task conditions in the top row, and memory paradigm in the bottom row. IF - impression formation; Memory - memory set.

UG - undergraduates

B/A - time of stereotype presentation, before or after presentation of to-be-remembered material

and Hastie presented evidence that the divergent results in the literature depended to some extent on whether the subjects received memory-set or impression-formation instructions. Hasher and Griffin (1978) pointed out the importance of different circumstances of retrieval. More recent formulations are more sophisticated, taking account of such variables as task goals, subjects' perspective, and judgment factors. These formulations are considered here.

1) THE SRULL-HASTIE MODEL Although Hastie, Park and Weber (1984) do not present an explicit model of stereotypes and memory, they do argue for the "Srull-Hastie model", based on work by Srull (1981) and Hastie (1980). Fundamental to their perspective is the *inconsistency effect*. They hold that unexpected behavioural information will receive more elaborate processing, at least under impression-set, although arguably not memory-set, instructions. They reinterpret previous research in an attempt to show that the inconsistency effect underpins the literature. The discrepencies in the literature result, they claim, from the various task methodologies employed. For instance, Rothbart, Evans and Fulero (1979) found a consistency effect; but Hastie *et al* hold that this was an artifact of using inappropriate dependent measures. Srull (1981) claims that the Rothbart, Evans and Fulero results might be due to the fact that they considered stereotypes of groups rather than individuals. Similarly, he

attributes the consistency effect found by Cohen (1981) to response-bias.

This formulation is tied in to depth-of-processing models, particularly the work of Bradshaw and Anderson (1982); who presented subjects with lists of facts about historical figures, one of which was a target or to-be-remembered fact, and the others "context" facts. There were four experimental conditions. In one condition, context facts were *causes of* target facts; in another *effects of* target facts; in the third unrelated. In the fourth condition, no context facts were presented. They found that facts related to context (whether causes or effects) were best recalled. They claim that this was because related facts received more elaborate processing than unrelated facts; and Hastie *et al* hold that items which receive more elaborate processing - in the present instance, stereotype-inconsistent items - will be best recalled.

This highlights the problem with the "Srull-Hastie model": they claim that the inconsistency effect underlies all the results in the area, but try to gain support for this contention from the consistency effect reported by Bradshaw and Anderson. The arguments presented in support of the model depend largely upon *post hoc* reasoning: it is unconvincing, for example, to explain away results which are at odds with the model by reference to methodological factors, (e.g., using groups rather than individuals as target

stimuli,) which are not explicitly set out in the model in the first place.

The issue is further complicated by the results of Bodenhausen (1988) who finds that consistent information receives more elaborate processing than inconsistent information at the encoding stage, and hence, according to the Srull-Hastie model, should have a memory advantage. Sentis and Burnstein (1979) find that schema-consistent information is easily chunked and stored in memory, whereas pieces of inconsistent information must be stored on an individual basis, and will therefore, being unrelated to other items in memory, be harder to access.

Conclusions The Srull-Hastie model, at least as presented here, appears unconvincing. Its predictions contradict many of the findings in the stereotyping literature, and these contradictions are explained away by *post hoc* reasoning. Also, the model is founded to some degree upon findings (Bradshaw and Anderson, 1982) which it would not itself have predicted. However, the model deserves credit insofar as it takes into account factors more subtle than merely the consistency or relevance of the information presented. More refined versions of the model may yet prove to have broader predictive powers than that presented by Hastie *et al* (1984).

2) THE WYER AND GORDON MODEL Wyer and Gordon (1982) present a model which is nominally an extension of the Srull-Hastie model. It discriminates between memory for individuals and for groups, and between

impression-formation and memory-set task characteristics. Under memory-set conditions, they hold recall to be governed largely by distinctiveness. Stereotype-consistent information, being less distinctive than -inconsistent information, is likely to be recalled less well. Under person-impression conditions, however, a trait-based representation of the target is formed by which the behaviours in target material are linked network-fashion to traits which they exemplify. Hence, trait-based descriptors act as retrieval cues for behaviours associated with them, causing a consistency effect.

Where subjects are required to form an impression of a group rather than an individual, Wyer and Gordon claim that a consistency effect is likely; except in cases where extra cognitive effort is required for processing of target information. This may occur where the behaviour is itself inconsistent. This argument is somewhat *post hoc* - one might ask how to predict under what circumstances extra cognitive effort will be required for processing - and appears to run counter to the conclusion of Rothbart, Evans and Fulero (1979): although see Wyer and Gordon (1982, pp 160 - 161) for a full discussion of this apparent conflict.

The model was expanded by Wyer and Gordon (1984) and Wyer, Bodenhausen and Srull (1984). Wyer and Gordon (1984) present a review of the social memory literature, from which they draw three general

conclusions. All of these apply exclusively to impression-set conditions. They are:

i) subjects encode behaviours in terms of the traits they exemplify

ii) when subjects have prior knowledge of a target's traits, they generate a *general impression* of the target. Once this is formed, inconsistent behaviours will be processed more extensively (Hastie, 1980). This occurs only with evaluatively, and not descriptively, inconsistent behaviours

iii) when subjects have no trait-based expectancy, they do *not* immediately form an evaluative impression of the target, but spontaneously encode and organize behavioural information in terms of trait concepts.

The aim of the paper by Wyer *et al* (1984) was to explain the underlying basis of these effects. Their explanation consists basically of three postulates: first, associations between a target's traits and behaviours, or among the behaviours themselves, are formed as a result of thinking about them in relation to each other. Second, the likelihood of recalling a specific behaviour increases with the strength of its association with a trait-based concept of the target. Third, the likelihood of recalling a specific behaviour increases with the number of other behaviours with which it is directly associated. From this standpoint, they are able to generate algebraic formulations of the probability of recall or recognition of different types of information under different target conditions (i.e., an individual or a

group). While it is not the aim of the present review to expound upon each of these formulations and its theoretical justification, a few words about their model of processing consistent and inconsistent information would be appropriate here. Wyer *et al* find a recall advantage for behaviours which are evaluatively inconsistent with trait-based target expectancies. In many instances this will not occur however: the most obvious being when trait expectancies (stereotypes by any other name) are introduced only after receipt of to-be-remembered material. Where a subject's expectancy is based upon a general description of the group to which the target belongs, however, rather than a description of the target person him/herself, then a target person deviating from the stereotypic behaviour of that group will likely be labelled "exceptional", and his/her behaviour ignored when forming an impression of the group as a whole. Hence, exceptional behaviour may be remembered less well than the more stereotype-consistent behaviour of other group members.

Wyer *et al* also present evidence that recall and recognition paradigms may access different aspects of target material. However, due to the complexity of the area, and also the limited research on the topic, they do not attempt to incorporate this distinction as a central feature of the model.

Conclusions The great strength of this model is probably that it discriminates less between consistency and inconsistency than between the

distinctiveness of different pieces of information with regard to the stereotype which they exemplify. It is possible that the findings presented in Table 6.1 will one day be united by the expedient of drawing in a third factor which covaries with consistency and relevance to a stereotype. However, the difficulty of incorporating such findings as Hastie and Kumar (1979), Hashtroudie and Mutter (1980), and Hemsley and Marmurek (1982), into the model, make it as yet an unconvincing account of the literature. Expansions of the model by Wyer and Gordon (1984), and Wyer, Bodenhausen and Srull (1984), make this probably the most sophisticated and detailed account of social memory yet presented. The fact that these later formulations consider impression-formation to the almost complete exclusion of memory-set paradigms however greatly limits their range of explanation; making it less applicable to the eyewitness memory literature than it might have been.

3) THE STANGOR AND RUBLE MODEL Stangor and Ruble (1989) consider both memory-set and impression-formation paradigms. They argue that the controversy over the "consistency" vs. the "inconsistency" effect may reflect a failure to take into account the knowledge available to the perceiver as target information is processed. The stronger the perceiver's expectations, the more likely is expectancy-consistent information to be recalled. This is supported by an experiment they report, in which subjects received behavioural descriptions of

members of two college fraternities, one of which engaged in mostly extroverted behaviour, and the other in mostly introverted behaviour. Half of the subjects first saw an "expectancy-generating" presentation of thirty behaviours. Half received memory-set instructions, and half impression-formation instructions. Both instruction sets were found to improve recall.

The fewer expectations a subject has about a target person or group, the more elaborate processing the subject must use in order to form an impression. In laboratory-based studies, which often employ stimuli for which subjects will have few expectations, inconsistent behaviours are likely to receive more elaborate processing: this explains results like those of Hastie and Kumar (1979), Hashtroudie and Mutter (1980), and Hemsley and Marmurek (1982): who all found inconsistency effects for material consisting of randomly-generated traits applied to an invented target. Studies which employ stimuli for which expectancies are greater (e.g., Halpern, 1985; Koblinsky and Cruse, 1981; Martin and Halverson, 1983, who all considered gender stereotypes) are likely to show consistency effects. In real-world situations, Stangor and Ruble argue, the expectations of a perceiver are likely to be more well-developed than is usually the case in a laboratory experiment; again leading to a consistency effect.

This is reinforced by a review of gender stereotyping and memory presented by Ruble and Stangor

(1986), who find a marked consistency effect in the literature: gender stereotypes presumably having high expectancies associated with them. A similar conclusion was reached by Higgins and Bargh (1987) in a review of the social memory literature. Stangor and Ruble conclude that consistent information is likely to be facilitated in memory in all instances *except* when a subject is forming a first impression of a target.

Conclusion This is probably the most subtle formulation yet to be applied to the literature: differentiating as it does not only between the results of impression-formation and memory-set tasks, but between different *stages* of the impression-formation process. Conceivably, the model could be tied in with the depth-of-processing literature in the same way that the "Srull-Hastie model" is tied in by Hastie *et al* (1984). For example, Hamilton (1981) holds that schema-relevant information will be given more attention, and is more likely to be noticed, than schema-irrelevant information.

The Stangor and Ruble model is probably the best-supported model - and the one with most explanatory power - in the psychological literature at present. However, it is certainly not perfect. The volume of research and number of contradictory findings in this literature demand that much more work be carried out to test and refine the model.

It is evident that none of these models is entirely convincing; however, all probably have something to

recommend them. Whatever model may in future be developed to account for the divergent findings in the literature, it is likely to resemble recent models more than earlier formulations. It must take account of the way in which results vary with task demands, and particularly with impression-formation and memory-set demands. Quite possibly, it will also take account of different stages that subjects pass through during each of these tasks; a problem not neglected by Stangor and Ruble. It must also show an awareness of depth-of-processing effects: certainly, as Bodenhausen (1988), and Hastie *et al* (1984), point out, information which receives the most elaborate processing is likely to show a memory advantage. Something which is missing from at least the majority of models reviewed here is a means of predicting, a priori, which items of target material are likely to receive most elaborate processing. Finally, although the literature has hardly considered this at all, it is conceivable that different results may be achieved with different memory paradigms (Wyer *et al*, 1984): memory advantages may vary from free-recall to recognition paradigms.

Stereotypes Presented at Retrieval

Very little theoretical work has addressed the issue of the effect of stereotypic information presented during retrieval. As mentioned above, the majority of models that have been presented in the literature deal exclusively with stereotypes presented at encoding.

This unbalance threatens to create something of a vicious circle in the literature, by which models of the one sort of effect are tested by means of experiments investigating that sort of effect, which in turn lead to further refinement of the models: implying that ultimately the effect of stereotypes presented at retrieval may come to be ignored altogether.

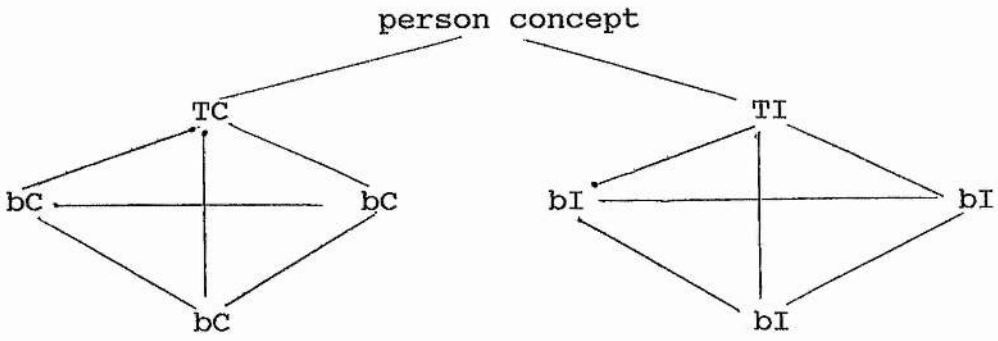
Investigations into the effects of stereotypes presented at retrieval (e.g., Snyder and Uranowitz, 1978c; Clark and Woll, 1981) have tended to formulate the problems to be investigated in terms of schema theory: the cognitve generalizations that are made about the characteristics of events, objects, or people. Such generalizations affect subsequent cognitive events associated with the objects of the schemata: i.e., information processing, cognitive integration, and organisation process.

It is not difficult to see how cognitive schemata might affect the retrieval of information from memory: even when schemata are activated *after* target information is encountered. Retrieval processes might be guided by schemata; hence the subject would be more likely to retrieve schema-consistent than inconsistent information. For instance, given the generic schema of a prospective purchaser of property, subjects remember from previously-presented target information that which is relevant to the schema (i.e., information which might affect the value of the property); whereas given the schema of a burglar,

subjects remember from the same target information that which is relevant to this schema (i.e., that which affects the ease of entering the house; the value of the contents, etc. - Anderson and Pichert, 1978). It is an explanation of this sort that underpins intepretation of Snyder and Uranowitz' (1978c) seminal "Betty K" experiment, and its subsequent replications (these experiments are discussed in some detail below).

Probably the best theoretical formulation of the effects of stereotypes presented at retrieval is however that offered by Wyer, Bodenhausen and Srull (1984). The experiment relevant here was one of a series of studies aimed at building and testing the social memory model discussed above; hence it did not provide the focus for the entire paper.

They present a model whereby trait information given at retrieval cannot lead to the reorganisation of that material around trait information. Thus subjects encode behaviours mentioned in to-be-remembered material in terms of "trait-behaviour clusters". The behaviours within each cluster will become more strongly associated than those across clusters. When trait information is provided, a central concept is activated, and the clusters become associated with it as shown overleaf, where T is trait, b is behaviour, C is trait-consistent, and I is trait-inconsistent:



Lines between letters represent interrelationships among the behaviours. The red line represents a stronger interrelationship; blue lines weaker interrelationships.

From this position, they predict a consistency effect for both free-recall and recognition tasks like that presented in the Snyder and Uranowitz experiment. Both predictions held up when tested. These are the same predictions as given by the theory that stereotypes act as retrieval schemata. Therefore, any successful replication of the Snyder and Uranowitz experiment, or unambiguous demonstration of a consistency effect where stereotypic information is presented at retrieval, will tend to support the model proposed by Wyer, Bodenhausen and Srull.

Methodological Aspects of Stereotyping Research

Stereotypes Presented at Encoding: Table 6.1 indicates that there is little variety in the methodology of those studies that have examined the effects of stereotypic information on memory. The majority tend to use the same paradigm: undergraduate subjects receive a piece of stereotypic information,

followed by a piece of to-be-remembered material (usually in the form of prose), under instructions either to try to remember as much information as possible, or to form an impression of the target(s) described. Subjects' memory for information from the original material is then tested, to see which type of information is best recalled: stereotype-consistent, stereotype-inconsistent, or stereotype-irrelevant information.

The results obtained vary far more greatly than the methodology, however. Table 6.1 details twenty-two studies in which stereotypic information was presented at encoding of to-be-remembered material. Of these, thirteen found a consistency effect, eight an inconsistency effect, and one considered only recall levels of relevant vs. irrelevant information (Bower, Black and Turner, 1979). Roberts (1985), in his review of the literature on the effects of subjective attitudes on memory, holds that little variation here is due to consistency at all. Many studies consider, as a side-issue, the related question of the relevance of information presented. Rothbart (1981; Rothbart, Evans and Fulero, 1979) finds that irrelevant and inconsistent information are recalled equally well, but less well than consistent information. All other studies of this effect find lower recall levels for irrelevant information than for consistent or inconsistent information. A number of attempts have been made to give a coherent account of the apparently

contradictory results of the studies reviewed in Table 6.1. These will be considered below.

The literature on the effects of stereotypic information presented at retrieval is far smaller, consisting as it does essentially of one classic experiment - Snyder and Uranowitz (1978c) - and three attempts to replicate it. This important area receives close attention later in the present chapter, and in Experiments 5 and 6.

Table 6.1 also shows that few recent studies have been reported. The emphasis in the literature has recently moved away from demonstrations of effects of stereotypic information on memory and towards the applications of theories which have developed from this work (see, e.g., Macrae, in press). However, it is hoped that the present chapter highlights a number of important issues that still require resolution, and potential real-world ramifications which this might have.

Stereotypes Presented at Retrieval: In many ways, it is the effect of stereotypic information presented at retrieval that is of most relevance to the present thesis. This area will be investigated in Experiments 5 - 7. The area has been much ignored: the majority of studies mentioned in Table 6.1 and above consider only the effect of stereotypes presented at encoding: hence the fact that most of the models reviewed take only this paradigm into account. This threatens to create a vicious circle in the literature: experiments that investigate one sort of effect generate models of

that effect, which in turn generate more experiments; and the effects of stereotypic information presented at retrieval are therefore ignored. This has two important ramifications: 1) investigation into the effects of stereotypic information presented at retrieval is proceeding at a crawl. A number of important questions in the area need to be addressed; and have done for many years 2) theories of social memory based solely on results achieved in one paradigm may be proven inadequate by successful work in the complementary paradigm.

There is reason to believe that stereotypes may affect memory when presented after to-be-remembered material. Researchers in related fields have found that information presented after target material can affect recall or recognition of that material. Arkes and Harkness (1980) found that subjects given a diagnosis of a medical condition (Down's syndrome) after receipt of information about symptomatology were better able than those not given such information to recognise consistent symptoms after a twelve day retention interval, and correctly to reject new, non-presented symptoms. Dooling and Christiaansen (1977) found that subjects told after reading a biographical passage that it related to the life of a famous person later falsely "recognised" sentences which, while not present in the original passage, were consistent with the life of the famous person. Lingle, Geva, Ostrom, Leippe and Baumgardner (1979) found that making a judgment about a target person on the basis of trait

information previously received about that person influenced the traits later recalled. Ross, McFarland and Fletcher (1981) manipulated subjects' attitudes to an issue, and then asked them to recall past behaviours relevant to the newly-formed attitude; finding that items recalled differed with attitude. This is a particularly interesting study in that it shows the effects not to be limited to the recognition paradigm but also to hold in the recall paradigm.

Experiments 2 and 3 in the present thesis find no effect of stereotype presentation time: biassing effects were found both when stereotypic information was presented at encoding and when it was presented at retrieval. The same effect may be expected to hold in an eyewitnessing situation: a piece of stereotypic information encountered after witnessing a crime (e.g., during discussions with friends and relatives, or during the police interview) may bias memory for the crime.

There are other reasons to predict an effect of stereotypes presented at retrieval. Generic knowledge structures or schemata may guide memory by providing retrieval cues (Snyder and Uranowitz, 1978c; Lutz, 1983). Clearly this is an important area for consideration in the present thesis. The studies in this area are reviewed below.

ROTHBART, EVANS AND FULERO (1979): These researchers presented subjects with a list of fifty behaviour descriptions. Each behaviour was associated with one

member of a group of fifty men. Of the fifty items, seventeen were intelligent behaviours, seventeen friendly behaviours, three unintelligent behaviours, three unfriendly behaviours, and ten were unrelated to intelligence or friendliness. For half the subjects, the group of men was labelled "intelligent", and for half "friendly". Half the subjects in each group received this label at encoding, and half at retrieval. Behaviours consistent with labels presented at encoding were better recalled and received higher frequency estimates than inconsistent behaviours. No converse effects were discovered: presentation of a label at retrieval was not found to affect memory in any way. Taken alone, this study implies that, in the type of situation of interest here, a piece of stereotyping information accidentally dropped into conversation during a police interview, say, would not be expected to affect the memory of a witness to a crime.

SNYDER AND URANOWITZ (1978C): Snyder and Uranowitz presented subjects with a biography of a fictional character named Betty K. The biography was constructed in such a way that equal numbers of facts indicated that Betty was heterosexual or homosexual. After reading the biography, some subjects learned that Betty was currently living a homosexual lifestyle ("lesbian-label condition"), and some that she was living a heterosexual lifestyle ("heterosexual-label condition"). Still others were given no such

information ("no-label condition"). These pieces of information were intended to stereotype Betty as lesbian or heterosexual, or to give no stereotypic information about her.

The subjects were then administered a forced-choice recognition task consisting of thirty-six items assessing the impact of stereotypic information on their memory for the biography. Each item allowed subjects to choose one of a number of responses: which Snyder and Uranowitz called "alternatives". Some alternatives were lesbian-label consistent; some heterosexual-label consistent; others neutral with respect to sexual-orientation label. All items had previously been rated by a panel of judges for how well it would test a subject's beliefs about the sexual orientation of a person. Items judged to test beliefs about sexual orientation were termed "critical" items. The questionnaire contained seventeen such items.

Subjects who believed that Betty was currently living a lesbian lifestyle remembered (or, in Snyder and Uranowitz' term, "reconstructed") the events of her life in such a way as to reflect stereotyped beliefs about lesbians to a greater extent than did those who believed that she was currently living a heterosexual lifestyle. The responses of subjects in the heterosexual-label condition did not differ from those of subjects in the no-label condition. Presumably this indicates that to label a person

"heterosexual" is equivalent to offering no stereotypic information about them.

Snyder and Uranowitz went on to test whether the results obtained were mediated by differential accuracy of memory in the two conditions (i.e., did receipt of a stereotype label make the subject more likely to choose a correct, label-consistent alternative?) or by differential error (i.e., did receipt of a stereotype label make the subject more likely to choose an incorrect but label-consistent alternative?) They found that subjects in the lesbian-label condition were significantly more likely to make lesbian-label consistent errors than those in the heterosexual-label condition. The converse was also true: Subjects in the heterosexual-label condition were more likely to make heterosexual-label consistent errors than subjects in the lesbian-label condition.

As far as accuracy measures went, for subjects in the heterosexual-label condition, a greater proportion of correct, heterosexual-label consistent answers was given. The converse was not found to be the case: receipt of a lesbian label did not increase accuracy of memory for correct lesbian alternatives.

Snyder and Uranowitz conclude that: "Our empirical research suggests that current beliefs can and do exert a powerful channeling effect in attempts to remember the past" (p 948). Although they do not explicitly state as much, this has been interpreted to mean that stereotypic information can act as a

retrieval schema for previously-learned material (despite the fact that receipt of a lesbian label did not appear to aid memory for lesbian items in the original study - see, e.g., Bellezza and Bower, 1981; Clark and Woll, 1981). Were this true, it would have obvious potential to aid investigation into eyewitness memory.

However, the literature contains two important failures to replicate the Snyder and Uranowitz study.

CLARK AND WOLL (1981): Clark and Woll point out a number of methodological weaknesses in the original study. Most relate to the items used to test memory, which appear to have been poorly thought-out. The number of items to which the correct response was a lesbian alternative was greater than the number of items to which the correct response was a heterosexual alternative. Also, the number of lesbian alternatives in a given item was sometimes greater than the number of heterosexual alternatives, and vice versa. Sometimes there were more neutral alternatives than heterosexual or lesbian alternatives. Obviously, predomination of one type of alternative would tend to bias responses to that item. With all of these factors balanced in their own study, Clark and Woll found no effect of stereotype label on memory.

Concluding that the failure of this conceptual replication might be due more to changes made to the design than to non-existence of the effect, Clark and Woll then ran a direct replication of the original

study, using all the original materials. Once again, no effect was found: results in the three conditions did not differ significantly. Although the lesbian- and no-label conditions produced much the same results as the corresponding conditions in the Snyder and Uranowitz study, the heterosexual-label condition did give different results, and these were the results primarily responsible for the failure of the replication. The experiment also included an "encoding" condition, whereby subjects were presented with the sexual-orientation label before reading the biography. Given the results from similar paradigms discussed above, one must expect some memory bias to result from this procedure; although whether one would do better to predict a consistency or an inconsistency effect is unclear. Neither effect was however discovered. Clark and Woll conclude that:

> "Sex role stereotypes may not have as powerful an effect on reconstructive memory as Snyder and Uranowitz suggested, or at least that their paradigm may not be adequate for uncovering such an effect." (Clark and Woll, 1981, p 1071)

Conceivably, Clark and Woll mean by this that the Snyder and Uranowitz paradigm is inadequate for uncovering the effect in question *when the experiment is designed more rigorously*. The original results may have been an artifact of methodological failings rather than the consequence of the action of sex-role stereotypes on memory. Their results appear to indicate that stereotypic information presented at retrieval may have no effect upon memory for to-be-remembered material.

This is of course possible. However, another of their findings casts doubt over the validity of their own design. As mentioned above, when they presented stereotypic information at encoding, they received neither a consistency nor an inconsistency effect. The effects gained under this paradigm differ considerably (see above) - but it is certainly unusual to achieve no results at all. One might therefore be tempted to conclude that some experimental artifact confounded their results. Without more details of their study (none have been forthcoming despite letters requesting them), it is impossible to speculate further about this.

BELLEZZA AND BOWER (1981): Bellezza and Bower also attempted a conceptual replication of the Snyder and Uranowitz study. Their hypothesis was that the original results were caused by "response-bias". In other words, they presumed that when a subject was presented on the forced-choice recognition task with an item the correct answer to which s/he did not know, then s/he would tend to guess: the direction of this guess biassed by the subject's knowledge of Betty's sexual orientation. Thus, if a subject did not know the right answer to a given item, but did remember that Betty was, say, a lesbian, then s/he would be likely to choose the lesbian alternative.

Bellezza and Bower used no "control" (i.e., "no-label") condition as used by Snyder and Uranowitz: presumably because, in the original study, the results

obtained in this condition did not differ significantly from those obtained in the heterosexual-label condition.

As an addition to the original procedure, Bellezza and Bower varied the type of distractor presented with correct alternatives in the forced-choice recognition task, i.e., so that a correct lesbian alternative might be paired in one condition with a lesbian distractor, and in another with a heterosexual distractor.

Two experiments were run with these materials. Experiment 1 showed no better memory for lesbian material in subjects given a lesbian label than in subjects given a hetereosexual label; and no better memory for heterosexual material in subjects given a heterosexual label.

> "...[there was] no greater memory sensitivity to lesbian information for subjects who were presented a lesbian label for Betty K than for subjects presented a heterosexual label. Similarly the heterosexual-label subjects showed no memory advantage when tested on heterosexual information. These results do not confirm the memory mechanisms proposed by Snyder and Uranowitz." (Bellezza and Bower, 1981, p 862)

Using a signal-detection paradigm, they did however find evidence of some response-bias in their subjects. There was not however strong support for either hypothesis: the retrieval-schema hypothesis, or their own response-bias hypothesis.

Experiment 2 was a replication of Experiment 1, except that the salience of labelling was increased by presenting subjects with four questions concerning

Betty's sexual orientation before completion of the task.

This time, greater response-bias was found towards lesbian alternatives when subjects were told Betty was lesbian than when they were told she was heterosexual. It was concluded that the original results were caused by response-bias, and not the mechanism originally proposed.

LUTZ (1983): Lutz essentially ran a direct replication of the original Snyder and Uranowitz study, but included a recall condition as a check for the recognition procedure. If the presentation of stereotypic information had an effect in the recognition procedure but not in the recall procedure, Lutz argues, then it might be possible that some bias existed in the recognition task itself.

After reading the biography, subjects were presented with stereotype labels either immediately or after a delay of one week. All memory tests took place one week after receipt of the biography. Recall data showed that delay reduced the effectiveness of the label in biassing a subject's memory, although recognition data was mixed. The essential finding of Snyder and Uranowitz, however (i.e., that a stereotype label presented at retrieval can distort memory for relevant biographical material) was replicated. There is however an important drawback associated with this study: Lutz employed the *original* materials, as used by Snyder and Uranowitz, and, as Clark and Woll (1981)

point out, these were not of the highest quality. A similar study using the materials generated by Clark and Woll or by Bellezza and Bower would have been of more value.

ONE MORE PIECE OF WORK: A minor contribution to this debate has been made by an undergraduate at Aberdeen University (Shepherd, 1991, personal communication). In this truncated version of the Snyder and Uranowitz study, the biography of a male was created and only the homo- and heterosexual label conditions run (as in Bellezza and Bower, 1981). Seventeen critical items were used; pre-rated as implying homosexuality or heterosexuality. The results achieved fairly well replicated those of Snyder and Uranowitz. Measures were taken of the accuracy and error rates found in each condition. Of these measures, only the rate of correct recognition for homosexual alternatives was unaffected by the presentation of a stereotype label. Strangely, this is precisely the measure that one would expect to be affected were Snyder and Uranowitz' explanation of their results correct, i.e., were stereotype labels acting as retrieval schemata.

Unfortunately, it has proven impossible to obtain any more information regarding this study. Although it must therefore be taken with caution, it does provide evidence which, if not exactly supporting Snyder and Uranowitz' contention, does appear to counter Bellezza and Bower's claims.

COMMENT: While the Clark and Woll critique of Snyder and Uranowitz' original experimental materials is well founded, it is not easy to explain their results. Certainly, the results of their first experiment appear to indicate that the original results were caused by poor testing procedure rather than by the explicit action of stereotypic information presented at retrieval; but this explanation is cast into doubt when one considers that their *direct* replication of the original study also failed to find the same results as Snyder and Uranowitz. Unless one argues that Snyder and Uranowitz included some other variation in their experiment which they failed to report, this does appear to make Clark and Woll's explanation appear tenuous.

Rather more interesting is the claim made by Bellezza and Bower that the original results were caused by response-bias. It is doubtful however that this will prove to be the whole story. Firstly, in Experiment 1, Bellezza and Bower found no response-bias when subjects were given a heterosexual label; although there was evidence of this in Experiment 2. It is hard to see how this result can be reconciled with their claim that the Snyder and Uranowitz results were caused entirely by response-bias. Secondly, in the original paper, Snyder and Uranowitz anticipated the kind of objection put forward by Bellezza and Bower. They employed a "fabrication group", whose job it was to invent answers to the forced-choice recognition task, based solely on the sexual label -

lesbian or heterosexual - without having read the Betty K biography. Since this group showed no evidence of response-bias, so Snyder and Uranowitz claim, neither should the "reconstruction" groups. Bellezza and Bower counter this argument by saying that Snyder and Uranowitz' fabrication group was probably just choosing its responses at random.

This is not convincing. It relies upon the implicit claim that the fabrication group received no information upon which to base their responses. (Otherwise, it is hard to see how Bellezza and Bower could claim that the responses were pure guesswork.) This is however untrue: the members of the fabrication group were told that Betty K was homo- or heterosexual. This is a minimal amount of information indeed; but it is precisely the same amount of information that Bellezza and Bower claim to have produced response-bias in their reconstruction group. Of course, they had also presented their subjects with the Betty K biography beforehand, but it is on the sexual-orientation label only - not on the biographical information - that they blamed the mooted response-bias in their subjects. One suspects that they are trying to have their cake and eat it.

This is not to come out wholeheartedly on the side of Snyder and Uranowitz, however. A look at the pattern of results that they achieved (discussed above) renders their explanation less than wholly convincing. After all, it is a strange kind of

retrieval schema the results of which are mediated to such a great extent by error.

One must not forget however that the accuracy (by which Snyder and Uranowitz mean the number of items correct, rather than the number correct divided by the total) of recognition of heterosexual information was increased by labelling Betty K a heterosexual. One would be more convinced, though, if the accuracy of recognition of lesbian information had also increased: as mentioned above, "heterosexuality" is probably not even a stereotype. If it is not a stereotype, one can hardly argue that the increased accuracy of recognition of heterosexual items in the experiment was due to the "heterosexual stereotype" acting as a retrieval schema.

More work is required in the area if any concrete conclusions are to be drawn. That the Snyder and Uranowitz experiment should prove so important is perhaps unsurprising given the relative simplicity and directness of its design. Therefore it would appear that the way forward in this area is to attempt to resolve the confusion surrounding this experiment. It is also important for the present thesis to attempt to resolve this question. The eyewitnessing situation is a real-world instance which offers great scope for the action of stereotypes presented at retrieval. After seeing a crime, the eyewitness may be exposed to stereotypic information at a number of stages: e.g., during conversations about the incident with friends and relatives, during the police interview, etc. If

stereotypes presented at retrieval can be shown to have effects upon memory for target information, then the potential for distortion of the memory trace in the eyewitness situation is very great. Indeed it could be argued that, in this case, stereotypic information has more opportunity to bias memory at retrieval than at encoding. This is the area which is addressed in the following two chapters.

Summary and Conclusions

There are two questions concerned with the effects of stereotypes on memory; and these two questions can be approached in either of two ways: using the social or the cognitive paradigm. The literature on the effect of stereotypes presented at encoding is large and contradictory, partly because of the number of paradigms used: impression-formation vs. memory-set instructions, recall vs. recognition, and so on. The earlier, simpler models of these effects, such as the schema-consistency model proposed by Cohen (1977/81) appear unlikely ever to give satisfactory accounts of the literature. Later models appear more successful; and of these, the one proposed by Stangor and Ruble (1989) is the best supported. This model predicts a consistency effect in most paradigms.

The literature of the effects of stereotypes presented at retrieval is especially relevant to the present area, because so many effects are conceivable here, both positive and negative. This literature centres around the controversy engendered by the

classic paper by Snyder and Uranowitz (1978c), the results of which would not be predicted by most current models of the effects of stereotypes on memory (e.g., Hastie, 1980; Srull 1981; Bodenhausen 1988), but which would, interestingly, be predicted by that proposed by Wyer *et al* (1984). Snyder and Uranowitz' results appear to receive support from studies by Lutz (1983) and Shepherd (personal communication), although Clark and Woll and Bellezza and Bower (both 1981) failed to replicate the findings.

Were the Snyder and Uranowitz hypothesis to prove tenable, this would be a particularly important area for consideration in the present thesis. If not, the possible distorting effect of stereotypic information presented at retrieval is itself an area worthy of study. The following section considers the effects of social factors in the eyewitness memory situation, with emphasis on the way in which such factors may result in the production of distorted testimony. Experiments 4 and 5 then go on to study the effects of stereotypic information presented at the retrieval of target information.

Social Factors in the Police Interview

Every situation encountered by an eyewitness to a crime is a *social* situation. The crime itself, subsequent discussions with friends and relatives about the crime, the police interview, and possible eventual court appearances, are all situations of a social nature. Therefore it is to be expected that

social factors which have been shown to affect memory or judgment will affect the accuracy of the eyewitness' account. Examples of such factors are the attitudes or opinions of the eyewitness with respect to the type of crime in question; the attitudes or opinions of the persons with whom s/he discusses the witnessed crime; and any social stereotypic factors which may accidentally be dropped into conversations regarding the crime. Perhaps the most likely stage of the eyewitnessing situation for social factors to have an influence is during the police interview. Given the potentially great importance of such factors, it appears strange that so little research has addressed the area: the great majority of research into the police interview has been concerned with such issues as the development of strategies to increase the amount or accuracy of information recalled, (e.g., research on the cognitive interview by Geiselman, Fisher and their colleagues, reviewed in Chapter 10,) or specific means of dealing with special categories of witness (e.g., the child witness - Davies, Flin and Baxter, 1986; Flin, 1988; Flin and Tarrant, 1989; Friedemann and Morgan, 1985; Moston, 1988/89 - or the mentally handicapped - Tully and Cahill, 1986). In general, social factors have received only cursory attention, usually as a side-issue only tangentially related to the main thrust of a paper or experiment.

Mention is made of social factors in the police interview in relatively early papers by Firth (1975) and Goodsall (1974); however these papers are

essentially documents written from experience by police officers without the time or resources to put their observations to the test under laboratory conditions. A paper by Palumbo (1975) is in a similar position, although this paper, also written by a professional policeman and aimed at an readership of policemen, is precocious in that it stresses the great importance of such social factors as the conversational nature of the interview, and establishing a positive relationship with the interviewee, a decade before Conversation Management (reviewed in Chapter 10) became an important area of research in the interviewing literature.

More recently - perhaps partly due to recent developments in Conversation Management techniques, which stress the importance of such social factors as rapport building in the police interview - social influences have received somewhat more attention.

Perhaps the most important figure in this area of the literature is Eric Shepherd, who (like Palumbo) has stressed the importance of viewing the police interview as a conversation between officer and witness, in the course of which the two attempt mutually to establish the facts of a particular criminal incident. E Shepherd (1985) holds that it is the responsibility of the interviewing policeman to build what he calls a "psychological bridge" - i.e., a rapport or mutual understanding - between himself and the witness, and has developed a police training scheme which aims to teach officers how to build this

bridge during the course of an interview. Lack of social skills on the part of an interviewing officer can affect the accuracy and completeness of the account which s/he elicits from the witness. Particularly dangerous is the police officer's tendency to conceptualize the interview situation only in terms of his/her own specialist knowledge and ignore areas of knowledge which are inevitably also important in the interview situation, especially the area of *discourse* - a highly complex kind of event, which police officers are untrained in interpreting.

In a later paper, E Shepherd (1991) goes on to discuss how such poor interviewing procedure can lead to the negative outcome of interviewee resistance. Although it is not the purpose of the present section to discuss the phenomenon of interviewee resistance in any detail, an example might help to indicate the importance of social variables here. While some earlier papers, (e.g., Firth, 1975,) recommend that an interviewing officer try to appear highly knowledgeable about the case in hand (even to the point of trying to appear far more knowledgeable than s/he actually is,) E Shepherd holds that such an approach may in fact prove counter-productive. However hard one tries to appear knowledgeable, he argues, one cannot fool the witness for very long. Any mistakes made by the interviewing officer in the content of his/her assertions regarding a case will inevitably lead to suspicion on the part of the interviewee, loss of confidence in and respect for the

officer, and will tend to generate interviewee resistance. This is a good example of the influence of social factors in the eyewitness situation.

Roy (1991) briefly discusses the effects of the lack of training in interview techniques on police investigators. He holds that the standard training method (i.e., the recruit accompanies an experienced officer during that officer's interviews) may equip investigators to become "report-takers", but does not aid them to become "information-gatherers" - i.e., investigators able to elicit more information from a witness than that witness would otherwise have provided. Roy holds that the application of Conversation Management techniques would overcome a large proportion of the fundamental problems which currently plague the police interview (these are discussed in Chapter 10). Most of these, (e.g., a staccato style of questioning, and interrupting the witness before s/he has finished speaking,) are once again *socially*-based, and spring from an ignorance of discourse, as argued by E Shepherd (1991).

Perhaps the most relevant study to the present area is that presented by Mortimer (1991), who looks at the influence of police officers' cognitive schemata of criminal offences on their perception of those offences. Mortimer argues that police officers make sense of an offence by applying to it the schemata that they have derived from past experience of such cases, as well as other sources. Such schemata can prove useful tools for investigating officers if used

in a flexible manner, but they must not be allowed to control an officer's behaviour during an interview with an eyewitness. Mortimer reports three experiments which aim to study the contents of such schemata for three types of crime, (shoplifting, actual bodily harm, and rape,) and the relationship between the complexity of the "template" and the officer's background. She found very little commonality in officers' templates of these types of crime. This could imply either that officers are open-minded with respect to these cases, or that they are prone to what Mortimer terms "click-whirr" (relatively non-cognitively-driven) responding to cases, on the basis of very little information about a case. These "minimal templates (Mortimer, 1991, p 22) could act either beneficially or adversely in the investigation. They could have a positive effect where a skilled officer is trying to learn as much detail as possible about an idiosyncratic event; but could have a detrimental effect where a case is particularly complex and the officer has difficulty in cognitively coping with the wealth of detail associated with it. In this case, s/he might be "victim to problems of integrating information and missing out crucial detail and even 'missing the point'" (p 22). Additionally, the lack of specificity of these templates makes the handling of information problematic, because they provide no mental pointers to follow. An officer is much more likely to probe those areas that are already defined by the cognitive

template than those which are not so defined; hence the resultant account of the crime in question is likely to be distorted in the direction of the contents of the officer's template.

As mentioned above, it seems strange that the influence of such social-cognitive factors in the eyewitness situation has received so little attention. It is the aim of the following section of the thesis to help redress this balance, by studying the influence of social stereotypes on memory when presented during the retrieval of target information.

CHAPTER 7 - RECONSTRUCTION OF THINGS PAST: RETRIEVAL SCHEMATA AND THE AFTER-EFFECTS OF STEREOTYPING INFORMATION

> *"It's a poor sort of memory that only works backwards."* *Lewis Carroll*: Through the Looking-Glass

Introduction

The literature review in Chapter 6 has demonstrated the importance of attempting to isolate the effects of stereotypic information on eyewitness memory. Distorting effects are conceivable whether this stereotypic information precedes or follows the witnessed crime. While the effects are perhaps likely to be of greater magnitude in the former case, the latter is even more relevant to the eyewitness situation. Stereotypic information could be introduced at almost any time following the crime: e.g., when the witness is thinking over the details of the crime, discussing it with friends and relations, or during the police interview itself. There is also the intriguing possibility that stereotypic information presented at retrieval might act as a retrieval schema for stereotype-consistent information from the crime itself. These possibilities merit close consideration in the present thesis.

To this end, it was decided to run an experiment conceptually based on that of Snyder and Uranowitz (1978c), but testing subjects' memory with a task designed to eliminate response-bias. Such a procedure was already extant: that designed by McCloskey and Zaragoza (1985).

McCloskey and Zaragoza (1985): This paper reported an experimental procedure designed to eliminate the response-bias effects which the authors claimed to be at the root of the "misinformation effect" investigated by Elizabeth Loftus and her colleagues (e.g., Loftus and Palmer, 1974; Loftus 1977; Loftus 1979; Loftus, Schooler and Wagenaar, 1985).

In a typical example of these studies, subjects were exposed to a series of slides of e.g., a traffic accident. They then read a piece of prose purporting to be an account of the slide sequence. For the experimental ("misled") subjects, some peripheral detail was altered in the prose account (e.g., it was claimed that a yield sign was present when in fact a stop sign was present). For control subjects, the prose account did not mention the sign. When tested with a forced-choice recognition task for the slide sequence, subjects in the "misled" condition were found to perform consistently more poorly on this detail than those in the "control" condition.

For many years, the main controversy that this finding engendered was whether the misleading information *replaced* the original in memory, or rendered it *inaccessible*. McCloskey and Zaragoza, however, claimed that misleading information in fact had no effect on memory for the original event; that Loftus' finding was the result of reponse-bias. In a study where the original information was a stop sign and the misled item was a yield sign, McCloskey and

Zaragoza showed that subjects in the "misled" condition would be statistically more likely to choose the yield sign than the stop sign, regardless of the effect of misleading information on the memory trace.

McCloskey and Zaragoza eliminated response-bias using a modified testing procedure. This was a variation on the standard test whereby instead of offering in the recognition task a choice between the original information and misinformation, they offered a choice between the original item and a foil, i.e., a similar object which was not part of the original information or the misinformation. Subjects were not asked to choose between (say) a stop sign (shown in the sequence) and a yield sign (misled information), but between a stop sign and a roundabout sign (information that was neither shown nor misled). (In fact, this particular condition was not used by McCloskey and Zaragoza, but simply serves illustrative purposes here.) Under this procedure, where a subject could not remember the answer to an item, then s/he would have to guess. Misled subjects who did not remember the correct answer would be correct 50% of the time, as would control subjects who did not remember the correct answer. Thus response-bias would be eliminated. This remains the procedure of choice for eliminating response-bias in the post-event misinformation situation (see, e.g., Wright 1992).

Use of the Modified Procedure: The question of response-bias in Loftus' work is analagous to that in

Snyder and Uranowitz' study. In each case, post-event information leads subjects to produce distorted "memories" of the target material; and, in each case, subsequent researchers have claimed the effect to be caused by response-bias. Therefore it was decided to apply McCloskey and Zaragoza's modified procedure to a conceptual replication of the Snyder and Uranowitz study. If the original results were still to hold under modified procedure conditions, this would provide strong support for Snyder and Uranowitz' position: that stereotypic information can act as a retrieval schema. The reverse results, obviously, would provide support for Bellezza and Bower's response-bias position. Such results would imply a lack of real effect of stereotypic information presented at retrieval. The hypotheses generated by the rival theories are presented in more detail, below.

The important models of social memory proposed by Wyer, Bodenhausen and Srull (1984) and by Stangor and Ruble (1989) also predict a consistency effect where stereotypic information is presented after target material. If the results of the present experiment support the Snyder and Uranowitz hypothesis, they would also provide support for these two theories. Results which contradict the Snyder and Uranowitz hypothesis, however, would also count against these models.

The Free-Recall Paradigm: Srull (1984) suggests that there may be a systematic difference between the results produced in recall and recognition testing paradigms. Given that eyewitness memory is for the most part accessed under recall conditions (during, e.g., the police interview), it was important for the present experiment to include a free-recall condition for comparison with the recognition conditions reported. Such comparisons have rarely been made in the social memory literature (Srull, 1984).

Snyder and Uranowitz (1978a - Snyder 1991, personal communication), did, however, run a conceptual replication of the Betty K study in which a free-recall measure was used. Subjects were required to write an essay that included as many facts from the target biography as they could recall. Judges classified these according to their best estimate of the subjects' beliefs about Betty's sexual orientation. Two classifications were discriminable: subjects who believed that Betty was homosexual, and those who believed that she was heterosexual *or* had no belief about her sexuality (judges were unable to discriminate between these two conditions). In another study, Snyder and Uranowitz (1978b) considered interpretive activities in reconstruction. Subjects in this study were required to answer questions about factual events in Betty K's life, and to *interpret* what meaning the events had for understanding Betty K. Interpretations were found to reflect current beliefs about Betty's lifestyle. This occurred not only for

events that fit with current beliefs (e.g., a subject who remembered that a supposedly lesbian Betty never had a steady boyfriend in high school might see this as an early manifestation of her lack of interest in men,) but also, conversely, for events that did not so fit (e.g., if this subject remembered that Betty did go out on dates in high school, s/he might interpret this as an attempt to "pass" as heterosexual - Snyder, 1991, personal communication, and see also Snyder, 1981). These demonstrations appear to show that the basic Betty K finding is generalisable across paradigms (cf. Srull, 1984). None of these studies has yet been published, and details about them are unavailable; hence the evidence that they provide is somewhat weak.

Lutz (1983) is the only published study to apply a free-recall paradigm to the Betty K materials. He found a powerful effect of the "heterosexual" label - which led to subjects producing a more "heterosexual" protocol - but no significant effect of delay of testing time (testing after one or three weeks). This pattern of results closely parallels that found in the recognition conditions also included in his experiment, implying again that the information produced during free-recall is comparable to that produced during recognition.

The recall paradigm is of relatively greater relevance to the eyewitness memory situation: recognition memory, while a factor in such instances as the identification parade, will usually take a back

seat to recall memory, which the police interview is designed to probe. Hence, a second experiment is reported here which employs a free-recall rather than a forced-choice recognition task paradigm, in an attempt to test the cross-paradigm generalisability of the results achieved. These experiments are reported separately as Experiments a and b.

A Note on Methodology: Snyder and Uranowitz, in their original experiment, tested subjects after a three-week delay from receipt of the biographical material. In the present instance, only the nature of the effects are of interest; not their magnitude. With this in mind, in the interest of methodological simplicity, testing was performed immediately after receipt of stereotypic information. This brief interval was necessitated by the use of the free-recall condition. Pilot work showed that subjects asked to complete a free-recall task one week after receipt of the Betty K biography managed to produce only two or three items each (*pace* Lutz 1983). Since such poor results would necessarily obscure the relatively small expected effects of stereotypes on memory, it was thought preferable to test memory immediately after receipt of the biographical information, producing meaningful data from all groups of subjects. Besides, there was no reason to suppose that a short delay would *increase* the effects of stereotypes on memory. If anything, it would lead to *smaller* effects, with the biographical information

still relatively fresh in the subjects' minds. Hence this was, if anything, a conservative departure from the original experimental design.

A Note on Terminology: - throughout the following sections, terms from signal detection theory are used. Hence "a piece of information correctly recognised as lesbian-stereotype consistent" is termed a "lesbian hit"; "a piece of information incorrectly 'recognised' as lesbian-stereotype consistent" is termed a "lesbian false alarm", and so on. Of course, in the modified test procedure, correct alternatives were presented with a neutral distractor rather than one consistent with the opposite stereotype. Hence, under modified procedure conditions, there was really no such things as "heterosexual-" or "lesbian false alarms". Nevertheless, these expressions are used with reference to the modified procedure, in the interests of consistency and ease of comprehension.

It is realised that terminology from signal detection theory is strictly inappropriate to the present study since no signal detection paradigm was used. It was decided however that in this case the advantages of brevity and comprehensibility outweighed those of fastidious terminology.

Throughout, the term "neutral" information is used to refer to information with no direct implications for the sexuality of Betty K.

The two experiments are reported separately below, as Experiments a and b.

Experiment 4a

Hypotheses: The hypotheses of the rival theories are presented below, with a brief explanation of the reasoning behind them.

RETRIEVAL SCHEMA THEORY

1. There will be more "lesbian hits" than "heterosexual hits" with a lesbian label under all test procedures. The retrieval schema theory holds that the effect by which the increase in hits under a lesbian label is mediated should be unaffected by the memory task used (see above).

2. There will be equal numbers of "lesbian hits" and "heterosexual hits" with a heterosexual label under all procedures. The hypothesis holds that a heterosexual label, being not a stereotypic label, should have no effect on memory performance, regardless of memory task used.

3. There may be more "neutral hits" with a lesbian label. This prediction is not part and parcel of the retrieval schema theory, but is consistent with the hypothesis and, if fulfilled, would support it rather than the response-bias theory. It is possible that the schema properties of a lesbian label are not restricted to information linked to that label; the retrieval of neutral information too might be facilitated.

4. At least as many "heterosexual false alarms" as "lesbian false alarms" under any condition. The response-bias theory would predict that there would be a greater number of lesbian- than heterosexual false alarms where a lesbian label is given because those are precisely the conditions under which response-bias is most likely to occur. The retrieval schema theory holds that under no condition should the number of "lesbian false alarms" be greater than the number of "heterosexual false alarms".

RESPONSE-BIAS THEORY

5. Fewer "lesbian hits" under the modified than the standard test procedure when a lesbian label is given. The response-bias theory holds that response-bias will act only under the standard test procedure conditions.

6. Fewer "lesbian false alarms" under modified than under standard test procedure. This is an extension of the previous prediction.

Methodology

Subjects: One hundred and twenty undergraduates were used in the experiment; sixty-five were female and fifty-five were male. Subjects were aged between seventeen and twenty-five years. Sixty received the standard test procedure, and sixty the modified test procedure. All were volunteers who took part for a small fee.

Materials: It was decided to base the experimental materials on those used by Bellezza and Bower, since they used a greater number of critical items than did

Snyder and Uranowitz. Before running the experiment however it seemed necessary to pilot the materials and ascertain their relevance to the present subject-population. The Bellezza and Bower experiment was run not only ten years earlier, but also in a different country. Obviously it was possible that stereotypes relevant to their subjects might not be relevant to the present subjects. One hundred Psychology undergraduates were asked to rate, on a nine-point scale, how far the alternatives for each critical item in the Bellezza and Bower study were consistent with a general lesbian or heterosexual stereotype that might be held by undergraduates at the University of St Andrews. No critical item was retained unless at least half of the subjects agreed that one alternative was significantly "lesbian stereotype consistent" and one significantly "heterosexual stereotype consistent". Only fourteen of the seventeen items used by Bellezza and Bower study fulfilled this criterion; hence only these fourteen items were used (see Appendix 17 for details of the construction of the recognition tasks). Bellezza and Bower's Biography A of Betty K was tailored to fit the retained items, such that the correct answer to seven of the critical items in the questionnaire was a lesbian alternative, and the correct answer to the other seven was a heterosexual alternative. Hence the materials used here were more balanced than those used by Snyder and Uranowitz. The resultant biography was 1021 words long (this is reproduced in Appendix 18).

Each forced-choice recognition task contained a total of twenty-six items. "Neutral" items were identical across procedures. In the standard procedure task however the correct alternative to any critical item was paired with a distractor that implied the opposite sexual orientation (i.e., a correct heterosexual alternative would always be coupled with an incorrect lesbian distractor and vice versa); whereas, in the modified task, correct alternatives were coupled with incorrect, *neutral* distractors. All such neutral distractors had been previously rated by one hundred subjects as being neutral with respect to sexual orientation. In some cases, these were the alternatives used in the Bellezza and Bower study that the subjects had previously rated as "neutral"; in other cases they had been generated specifically for the purpose and then tested on the subjects.

Each item in each forced-choice recognition task also contained a "No information provided" alternative, as used by Snyder and Uranowitz. It would have been equally possible to have included a "Don't know" alternative; this was however considered less appropriate. It was felt that to say "No information provided" is to make a fairly positive statement with regard to the prose account received, and that this was therefore the alternative most likely to lead to response-bias, i.e., to the subject actually choosing one of the other two alternatives. It was important to allow as much room as possible for

the action of response-bias in the standard test condition, since the other two conditions were designed to *reduce* response-bias; the aim of the experiment being to isolate the locus of Snyder and Uranowitz' effects.

Before completing their respective memory tasks, all subjects received a stereotyping paragraph describing Betty K's "current" lifestyle. This paragraph was identical with that used by Snyder and Uranowitz and by Bellezza and Bower. It stated that Betty was currently living either a stereotypically homo- or a stereotypically heterosexual lifestyle. Subjects were then asked briefly to answer four questions about Betty K. These questions were identical with those used by Bellezza and Bower in their Experiment 2 to increase the salience of the stereotype label received.

Procedure: The instructions used in the experiment were identical with those used by Snyder and Uranowitz (1978c) and then by Bellezza and Bower (1981). Subjects were allowed five minutes to read the target biography. They completed a short (five minute) unrelated filler task. They were then presented with a paragraph labelling Betty K as hetero- or homosexual, and with four brief questions to answer (see above). All subjects then completed their respective memory tasks.

Analysis: Dependant measures were the number of "lesbian" and "heterosexual" alternatives correctly chosen in each condition (i.e., standard or modified procedure, lesbian or heterosexual label), the number incorrectly chosen, and the number of correctly remembered neutral items.

Results

The raw data for Experiment 4a are presented in histogram form in Figures 7.1 - 7.4.

Analysis of variance was run which took into account the combined effects of recognition paradigm and stereotype label on the information generated in response to critical items. No main effect of experimental condition was discovered: $F(3,116) = 1.5280$ $p = 0.2110$. This meant that the total amount of information generated did not differ significantly across conditions. The amount of different types of information generated (e.g., "lesbian hit", "heterosexual false alarm") did differ significantly however: $F(3,348) = 1010.8050$ $p = 0.0000$, although this was unsurprising since it was evident from the raw data that the subjects had generated far more hits than false alarms: compare, e.g., Figures 7.1 and 7.3. A significant interaction was discovered between experimental condition and information generated: $F(9,348) = 3.7785$ $p = 0.0001$, i.e., the type of information generated did differ significantly across conditions (see Appendix 19).

FIGURE 7.1 - LESBIAN vs HETEROSEXUAL HITS, TWO TEST PROCEDURES, HET. LABEL

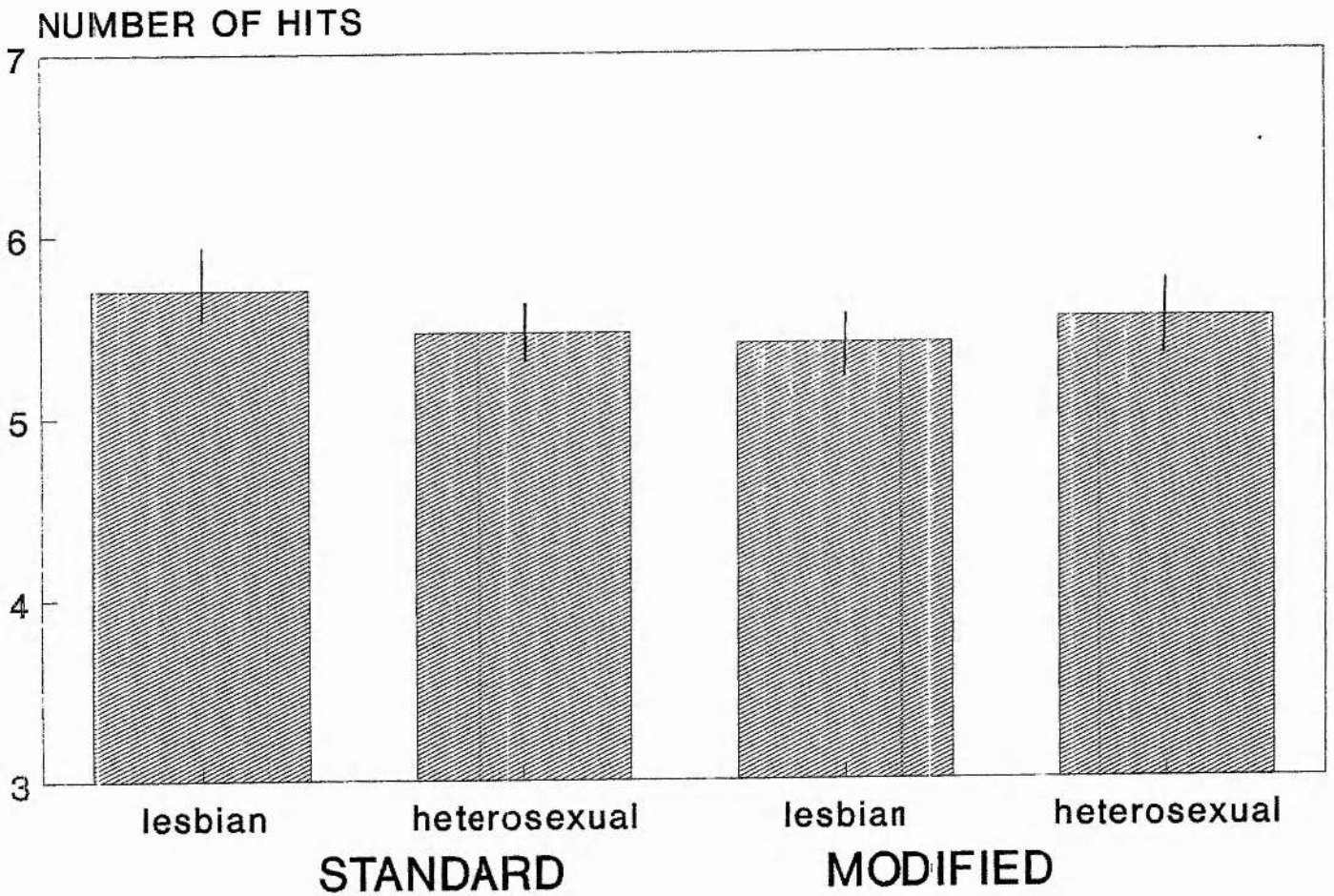


FIGURE 7.2 - LESBIAN vs HETEROSEXUAL HITS, TWO TEST PROCEDURES, LES. LABEL

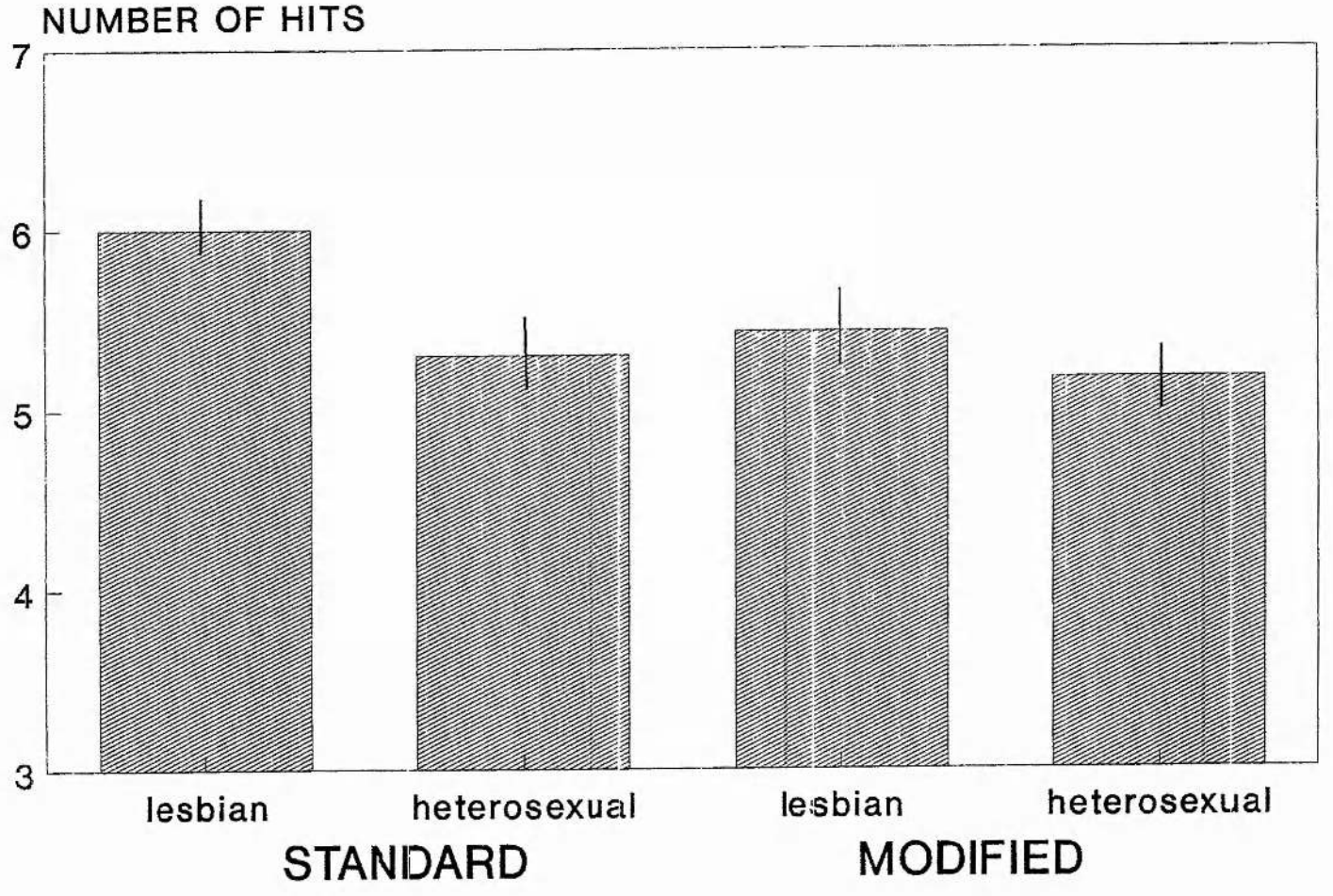


FIGURE 7.3 - LESBIAN vs HETEROSEXUAL FAs, TWO TEST PROCEDURES, HET. LABEL

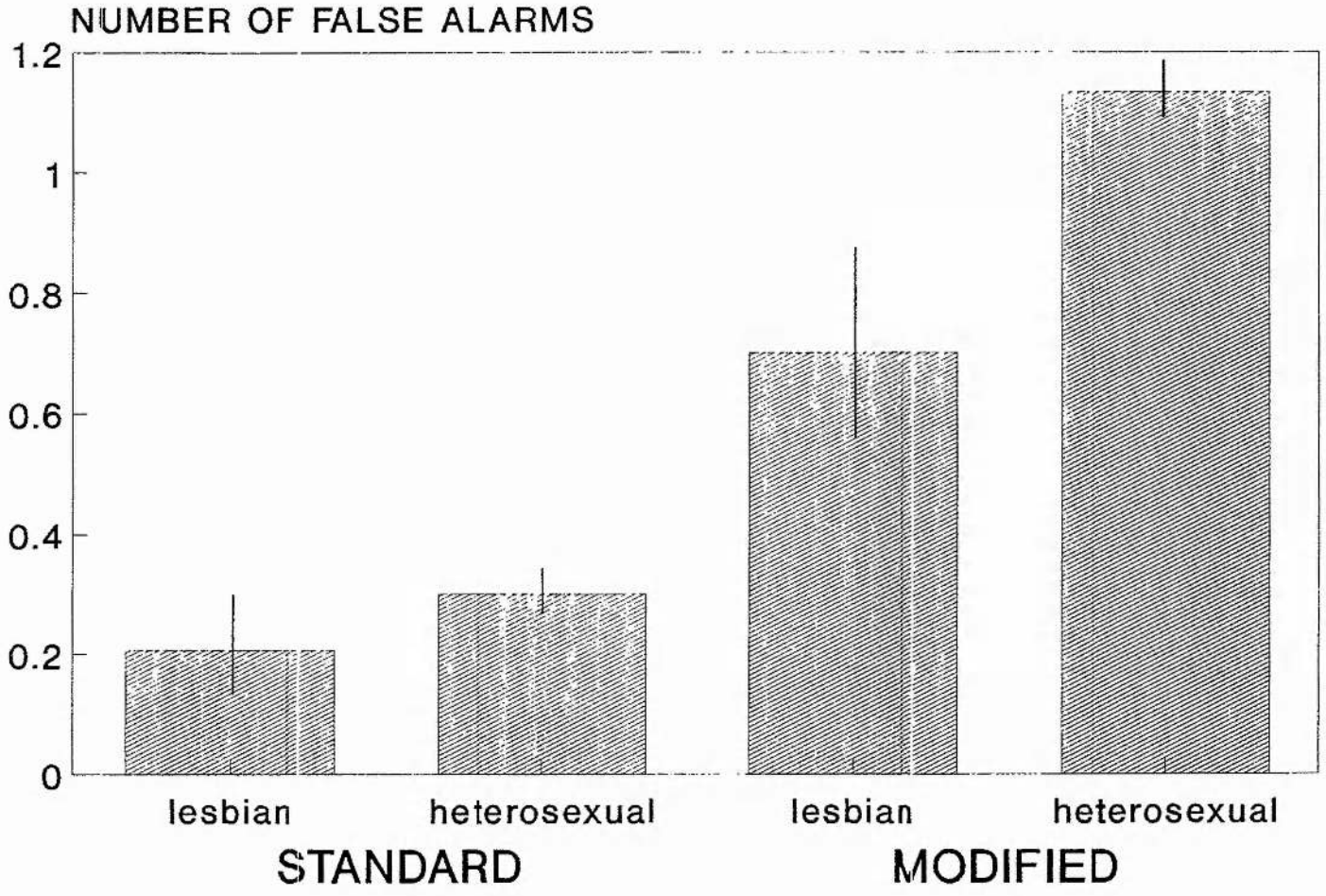
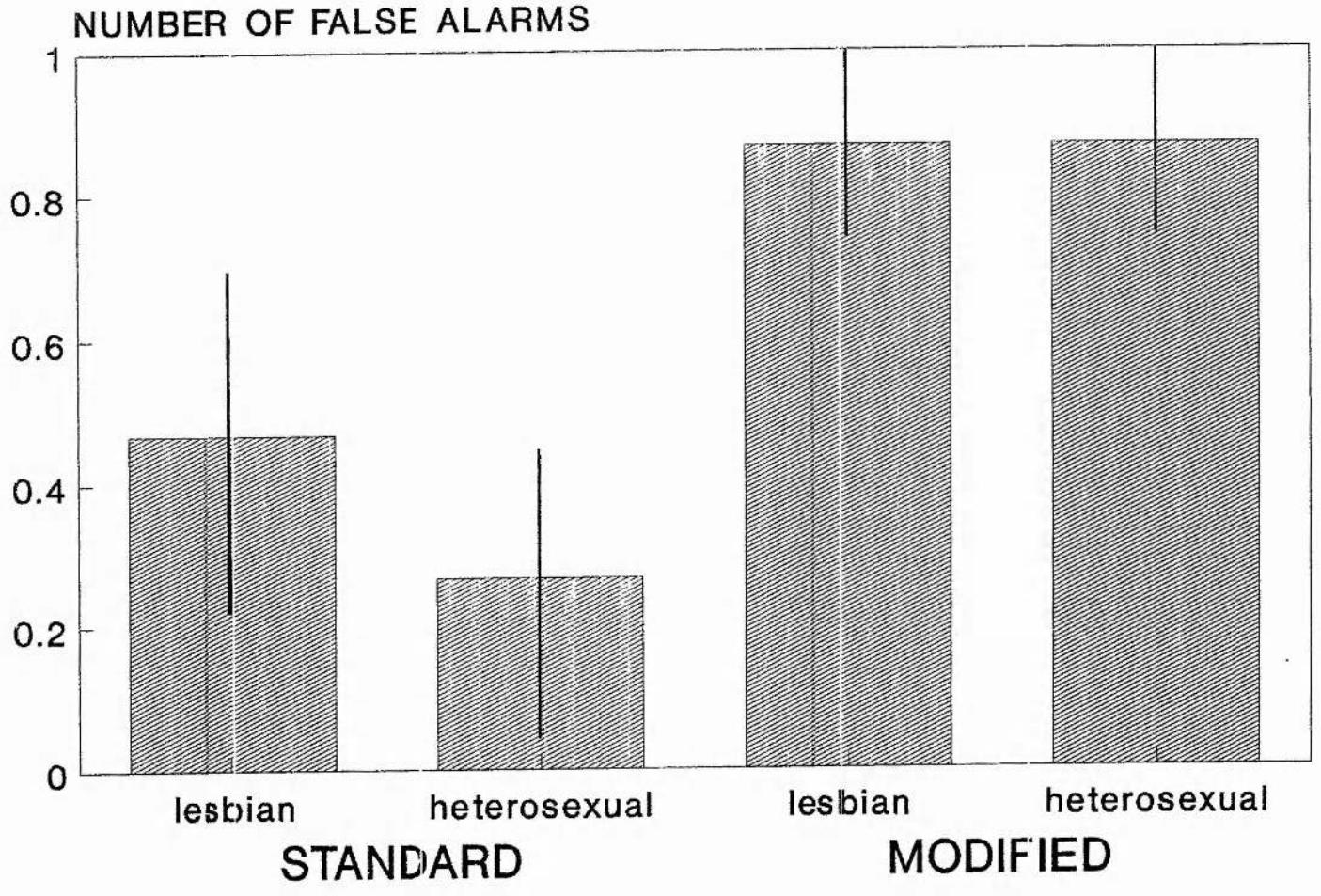


FIGURE 7.4 - LESBIAN vs HETEROSEXUAL FAs, TWO TEST PROCEDURES, LES. LABEL



Smaller analyses of variance attempted to break this effect down. Under a *heterosexual* label, the results of subjects under the standard and modified test procedure conditions did not differ significantly for correct information (or "hits"): $F(1,58) = 1.7617$ $p = 0.1896$ (see Appendix 20, Table 1). These data are shown in graph form in Figure 7.1. Subjects in each recognition paradigm, therefore, correctly recognised equal amounts of information. There was no significant difference between the numbers of lesbian and heterosexual hits made in either condition. This is consistent with the second hypothesis of the retrieval schema theory.

Where subjects received a *lesbian* label, there was again no significant difference between the two recognition conditions: $F(1,58) = 1.2406$ $p = 0.2699$, neither was there a significant interaction between recognition condition and information recognised: $F(1,58) = 1.5388$ $p = 0.2798$ (see Appendix 20, Table 2). Therefore, number of hits made under the lesbian label was unaffected by the test procedure employed. This result counts against the response-bias theory, but is consistent with the retrieval schema theory, which holds that the numbers of lesbian hits made will not differ with the testing procedure employed. These data are shown in Figures 7.1 and 7.2.

As for false alarms: under the modified procedure, subjects were significantly more likely to make false alarms: $F(1,58) = 28.9239$, $p = 0.0000$. This also counts against the response-bias theory, which

predeicted *fewer* false alarms to be made under the modified test procedure. There was also a trend towards more heterosexual than lesbian false alarms, although this was not significant: $F(1,58) = 3.088$ $p = 0.0841$ (see Appendix 21, Table 1). The fact that the numbers of heterosexual and lesbian false alarms made was not significantly different provides further support for the retrieval schema theory, which predicted that at least as many heterosexual false alarms as lesbian false alarms would be made under any condition. These data are presented in Figures 7.3 and 7.4.

The pattern of false alarms made under a lesbian label did however differ significantly across recognition conditions (see Figure 7.4). Under a lesbian stereotype label, subjects tested under the modified procedure made more false alarms than those tested under the standard procedure: $F(1,58) = 13.2571$ $p = 0.0006$. This result, casts some doubt upon the response-bias theory, which would predict that a testing procedure designed to eliminate response-bias should reduce the numbers of false alarms made with a lesbian label. However, the result does not have any direct ramifications for any of the hypotheses set out above. There was no significant interaction between recognition condition and type of false alarm made (i.e., between heterosexual and lesbian false alarms, $F(1,58) = 0.7484$ $p = 0.3906$: see Appendix 21, Table 2). This counts against the response-bias theory,

FIGURE 7.5 - NEUTRAL HITS
TWO TEST PROCEDURES, TWO LABELS

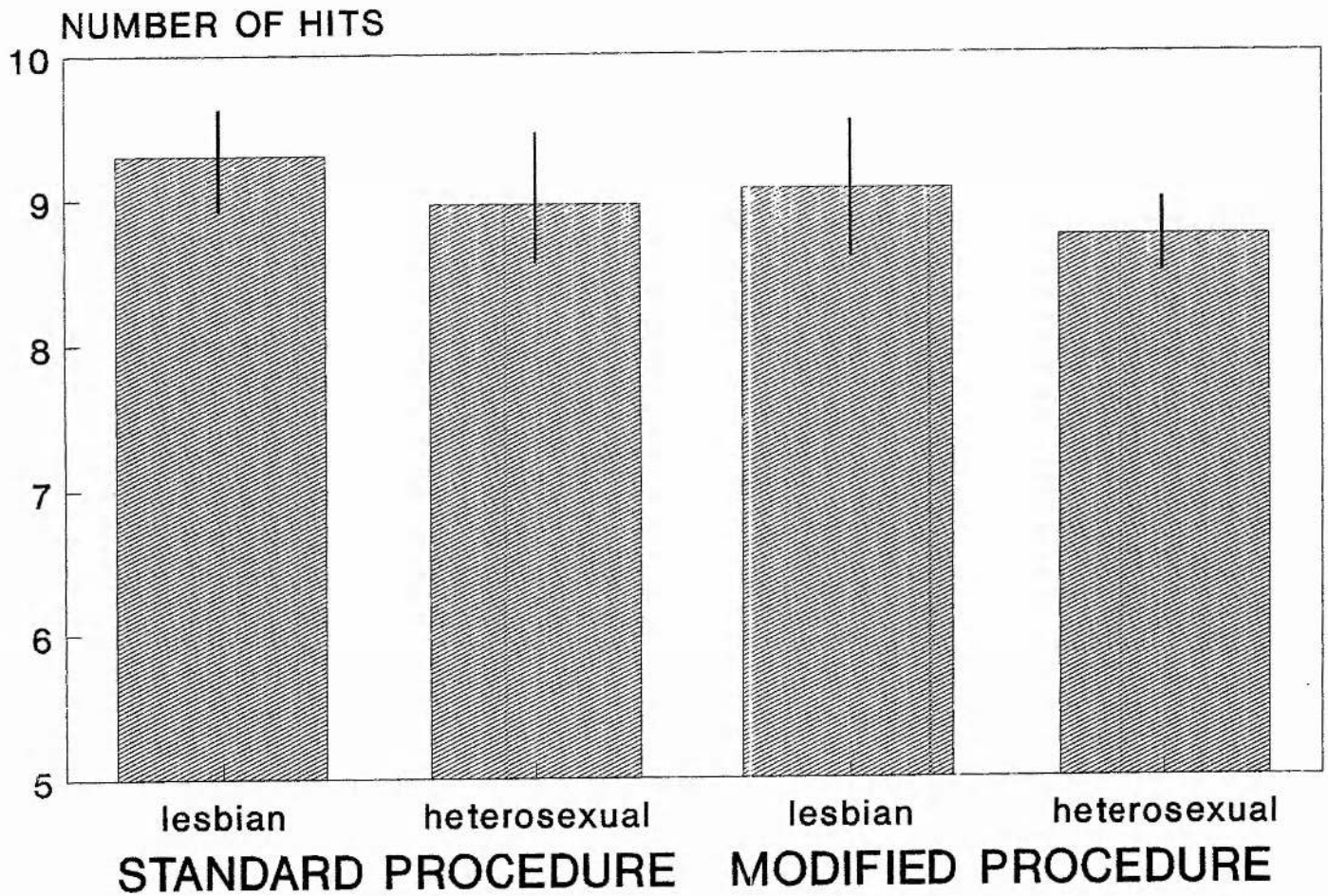
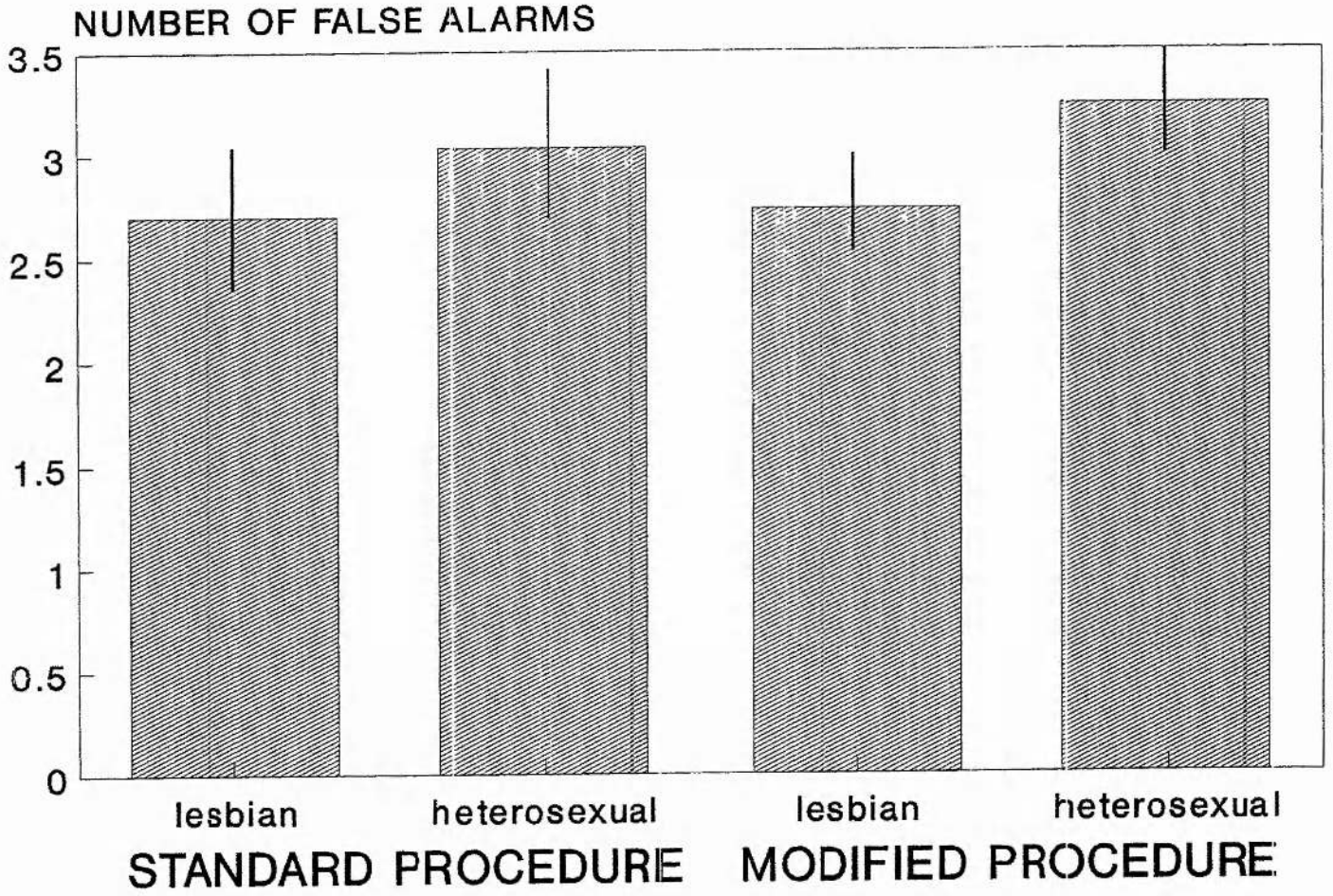


FIGURE 7.6 - NEUTRAL FALSE ALARMS TWO TEST PROCEDURES, TWO LABELS



which predicted that fewer lesbian false alarms would be made under modified than standard test procedures.

The amount of neutral information generated did not differ significantly across conditions (i.e., original and modified procedures x lesbian and heterosexual label) : $F(3,116) = 0.5567$ $p = 0.6447$ (see Appendix 22). The data for neutral hits and neutral false alarms made is rpesented in Figures 7.5 and 7.6. This disproves the third hypothesis of the retrieval schema theory; this was however not a hypothesis of central importance to the theory.

Discussion

Neither the retrieval schema nor the response-bias theory appears to give a fully convincing account of the findings of the present experiment. Of the two, the retrieval schema theory is certainly the more convincing, however. Neither of the predictions of the response-bias theory was supported.

The retrieval schema theory gives a superior account of the results. Four predictions were generated for this hypothesis, of which one was confirmed and one partially confirmed. It was predicted, in the second hypothesis, that equal numbers of lesbian and heterosexual hits would be generated under each recognition condition when a heterosexual label was given. This was confirmed: although it was found to hold only for the modified, but not for the standard, test procedure.

The fourth hpothesis held that there should be no difference in the numbers of heterosexual and lesbian false alarms made under any condition. Obviously, the finding that, overall, more heterosexual than lesbian false alarms were made would *not* have disconfirmed the hypothesis; therefore it was couched in the following terms: "At least as many heterosexual- as lesbian false alarms under any condition". Over all, there was no difference between numbers of lesbian and heterosexual false alarms produced; however, when a lesbian stereotype label was given, more lesbian false alarms *were* produced. This was caused by subjects in the modified test procedure making more lesbian false alarms than those in the standard test procedure. Hence the fourth hypothesis received only partial support.

Over all then the retrieval schema theory may have had more success in explaining the results; this success was not however complete. Two hypothess received no support.

Particularly damaging for the hypothesis was the finding that receipt of a lesbian stereotype label did not increase the number of lesbian hits made. It is hard to conceive of a retrieval schema that would act in such a manner; this was one of the major problems encountered by Snyder and Uranowitz' own interpretation of their results, outlined above.

Furthermore, the retrieval schema theory cannot explain why more lesbian- than heterosexual hits should have been made under standard test/heterosexual

label conditions. This finding counts against the hypothesis. Neither however can the response-bias theory account for this finding: Bellezza and Bower give no *a priori* reason to predict greater salience for lesbian information when a heterosexual label is received. Hence this finding does not appear to be explicable in terms of either hypothesis. Conceivably, lesbian-stereotype consistent material is more unexpected than heterosexual-stereotype consistent material in the context of a short biography, and for that reason more memorable, no matter what stereotype label might later be presented.

One other unexpected finding was that the subjects in the modified test procedure made more false alarms than subjects in the standard procedure, under both lesbian and heterosexual label conditions. This effect was especially pronounced for heterosexual false alarms. The retrieval schema theory would predict at least as many false alarms to be made under the modified as under the standard test procedure; whereas the response-bias theory would predict a greater number of false alarms to be made under the standard test procedure. Hence this finding tends to count against the response-bias theory. Nevertheless, it is a difficult result to explain. Tentatively, one might suggest an explanation: that some hitherto unsuspected artifact in the standard test procedure might act to reduce the number of errors made. However, this is necessarily a *post-hoc* conjecture, and cannot be made with any degree of certainty.

Summary: On the whole, the retrieval schema theory appears to offer a better explanation of the results achieved here than the response-bias theory. It is possible that at least some of the weaknesses of the hypothesis are explicable in terms of the paradigm employed: perhaps it was insufficiently sensitive to highlight all the effects taking place. This is quite plausible given that a forced-choice recognition task has of its nature a very rigid structure, and therefore might not allow subtle effects to be seen.

Although the predictions of the retrieval schema theory received only partial support in this experiment, it does appear a more convincing explanation of the results than the rival response-bias theory. If this experiment be taken as evidence that stereotypes can act as retrieval schemata, then it is evident that the retrieval schemata act only for material that is relevant to the stereotype. Subjects' recognition performance on neutral items did not vary with recognition paradigm or with label received.

Experiment 4b

This experiment tests subjects' free recall performance for the Betty K biography after receiving either the lesbian- or the heterosexual stereotype label. It is hoped that results from this paradigm might be used to resolve questions remaining after Experiment 4a, and also test in a more externally

valid manner the effects of stereotypes presented at the retrieval of target material. An eyewitness to a crime will far more likely be required to use a recall than a recognition strategy when interviewed by the police; hence it is important to study in the present thesis the implications of this memory paradigm.

Methodology

Subjects: Sixty subjects were used in the present experiment. Thirty were male and thirty female, all aged 17 - 25 years. All were volunteers who took part for a small fee.

Materials: The target biography, stereotyping paragraph, and four short questions used were identical with those used in Experiment 4a. All subjects were given a sheet of lined A4 paper on which to write down everything that they could recall from the Betty K biography. Subjects were under no compulsion to try to write a complete record of Betty's life. It was expected that this absence of compulsion would inspire subjects to record only those details of which they were fairly sure. Hence, any response-bias operative should be minimal, and any results achieved largely attributable to the action of a retrieval schema.

Procedure: The experimental procedure was almost identical with that of Experiment 4a, except that, instead of completing forced-choice recognition tasks,

subjects were asked to write down everything that they could recall from the target biography.

Scoring: Free-recall accounts were scored along the same dimensions as the forced-choice recognition tasks. This was done by a pair of judges who were blind to stereotype condition.

Prior to the start of the experiment, free-recall accounts of the Betty K biography were obtained from twenty volunteer subjects (Honours Psychology undergraduates). The judges were trained by scoring these accounts. Scoring was carried out separately, and correlations between the scores of the judges calculated ($r = 0.876$, $p < 0.01$). After the training period, the judges were allowed to discuss their scoring methods and resolve any disagreements they may have had concerning how particular responses should be scored. The judges were at all times required to concur with the scoring of the recognition tasks: e.g., an item scored as a "lesbian hit" on the recognition tasks was necessarily judged a "lesbian hit" in the recall task.

The same procedure as for the training period was followed for the scoring of the experiment proper. Judging was performed separately, but where there were disagreements these were settled by discussion. Results were analysed in the same way and along the same dimensions as those of the recognition task conditions.

Results

Analysis of variance performed on the entire data set indicated that the information generated differed significantly between the two free-recall groups: $F(1,58) = 5.2390$ $p = 0.0257$. Significantly more information was generated by subjects who received a lesbian label. The type of information generated also differed significantly: $F(3,174) = 165.4276$ $p = 0.0000$: subjects made far more hits than false alarms. There was a significant interaction between stereotype label received and information generated: $F(3,174) = 6.8146$ $p = 0.0002$. This meant that more lesbian hits were generated by subjects who received a lesbian label than by subjects who received a heterosexual label, although no more heterosexual hits were generated by subjects who received a heterosexual label than by subjects who received a lesbian label. There was no significant effect of stereotype label on the number of false alarms made: i.e., under a lesbian label, subjects were no more likely to "recall" a piece of incorrect information that was not in fact mentioned in the biography, although this misses significance by just seven ten-thousandths: $F(1/58) = 3.9835$ $p = 0.0507$ (see Appendix 23, Table 2). The response bias theory holds that lesbian false alarms are more likely under a lesbian label than heterosexual false alarms. Hence, this effect was broken down by analysis of variance of the type of false alarms made under the lesbian label. It was found that more lesbian false alarms were made under a

lesbian label than under a heterosexual label: $F(1/58) = 5.3506$ $p = 0.0243$ (see Appendix 23, Table 3). There was however no difference in the pattern of mistakes made under a heterosexual label: no more heterosexual false alarms were made than lesbian false alarms (totals were exactly equal).

A Neumann-Keuls test was run on these data and it was found that all significant results were caused by the comparison of "false alarms" with "hits" - and, as mentioned above, "hits" outnumbered "false alarms" in all conditions. Under a lesbian label, more lesbian hits were made than lesbian false alarms ($Q = 3.397$, $p < 0.05$), or than heterosexual false alarms ($Q = 3.7$, $p < 0.01$). Also under a lesbian label, there were significantly more heterosexual hits than lesbian false alarms or heterosexual false alarms ($Q = 2.967$, $p < 0.05$ in both cases). Finally, under a heterosexual label, there were significantly more heterosexual hits than lesbian false alarms or heterosexual false alarms ($Q = 3.467$, $p < 0.05$ in both cases).

There was no significant difference between the number of neutral hits made under the lesbian label and the number made under the heterosexual label: $F(1,58) = 1.1416$ $p = 0.2897$ (see Appendix 23, Table 4).

Discussion

It was hoped that any problems presented by the results from the recognition paradigms might be

resolved by reference to the results obtained under the free-recall paradigm. Although these data do not unequivocally support either side of the argument, again they do appear more consistent with Snyder and Uranowitz' retrieval schema theory than with Bellezza and Bower's response-bias theory.

Label received had a powerful effect on information generated. A greater number of lesbian hits was generated when a lesbian label was received, but the converse did not hold when a heterosexual label was received. This pattern of results is exactly in line with the predictions of the retrieval schema theory; being analagous to that found in the original Snyder and Uranowitz study.

The pattern of false alarms muddles the picture somewhat, however. More lesbian false alarms were made when a lesbian label was received. However, there was no difference in the pattern of false alarms made when a heterosexual label was received: under this condition, subjects were as likely to make lesbian false alarms as heterosexual false alarms. This finding favours the response-bias theory: subjects receiving a lesbian label appear to have been more likely to guess at pieces of information which were not actually in memory; these guesses being guided by the stereotype the subjects had in mind. Receipt of a heterosexual label, on the other hand, had no such effect, since, as argued above, to label a person a "heterosexual" is, more or less, to give no stereotyping information about that person.

This result may however simply reflect the stereotype of a lesbian is more developed than that of a heterosexual. Additionally, when these results are considered in absolute terms, the response-bias theory appears somewhat weaker. Although it was statistically significant, the absolute difference in the number of lesbian false alarms made by subjects across stereotype label conditions was a mere ten, i.e., one-third of one false alarm per subject. The absolute difference between the numbers of lesbian hits made however was thirty-eight, i.e., almost one and one-third hits per subject. Given that an increase of nearly one and a third items of correct information was achieved under the lesbian label at a cost of one third of an item of *in*correct information, one could be justified in claiming that a stereotype label does indeed act as a retrieval schema, and that if response-bias does occur, then that its effects appear comparatively small. This result is more in line with the retrieval schema theory than with the response-bias theory.

Again, it is obvious that while stereotypic information may act as a retrieval schema, this schema facilitates memory for relevant material only. Under free-recall, as under recognition conditions, stereotype label had no effect on memory for neutral items. This is unsurprising: although the third hypothesis presented above held that presentation of a lesbian stereotype might increase the number of

neutral hits made, this prediction was by no means central to the theory.

General Discussion

These experiments were intended to discover the locus of the effects found in the classic experiment by Snyder and Uranowitz (1978c), in which it was found that presentation of steretypic information after biographical material biassed memory for that material. This conclusion has been interpreted in subsequent literature as indicating that stereotypes can act as retrieval schemata. Bellezza and Bower (1981) claimed that they induced response-bias. The importance of this area to the eyewitness memory situation prompted an experiment to be conducted with the aim of choosing between the rival theories. This was done by including a recognition condition designed to eliminate response-bias, and a free-recall condition, the aim of which was to elucidate the effects observed in the recognition conditions.

Neither hypothesis was unequivocally supported. The pattern of results achieved was not fully explicable in terms either of response-bias or retrieval schemata. However, on the basis of the evidence presented here, the retrieval schema theory certainly appears the more compelling of the two. This is especially true of results reported in Experiment 4b, in which a free-recall rather than a recognition-task procedure was employed. As mentioned in the

Introduction, the free-recall paradigm is the more relavant to the eyewitness situation.

One line of evidence does count against the retrieval schema theory, however: in Experiment 4a, receipt of a lesbian label did not appear to facilitate recognition memory of lesbian-stereotype consistent material. Given the nature of the hypothesis, this was easily the most important of its predictions. This result was however discovered in Experiment 4b: the more relevant task to the present thesis. It could be argued the greater freedom implicit in a free-recall task allows schemata greater scope to act. This has powerful implications for police interviews: perhaps stereotypic information could be used as a last-ditch attempt retrieval mnemonic in particularly difficult situations. Such a technique would have to be kept as a last resort, however, given the potential distorting effects of stereotypic information on memory (see Chapter 6).

Nevertheless it cannot be argued that the results reoported here provide complete support for the retrieval schema theory. Consistency effects were discovered both by Snyder and Uranowitz (1978c) and Lutz (1983), using recognition paradigms, and materials similar to those employed here. This is an issue of much theoretical importance which deserves investigation. Srull (1984) argues that the quality of the memory trace for social information may alter with time, such that different memory-testing techniques may tend to give slightly different results

when presented after different intervals. This is the next area for investigation.

A more minor aim of these experiments was to test the theories of social memory proposed by Wyer, Bodenhausen and Srull (1984) and by Stangor and Ruble (1989). As mentioned in Chapter 6, these are currently probably the two best-supported theories in the social memory literature. Unfortunately it was not possible in the present experiment to test rival predictions of the two theories, since both hold that a consistency effect is likely where stereotypic information is presented after to-be-remembered material (except in the first stages of impression-formation, at which point the Stangor and Ruble model holds an inconsistency effect to be likely). Insofar as a consistency effect appears more prominent here, the present experiment tends to support both models.

It was hoped, using the present paradigms, to isolate the effects of stereotypes presented at the retrieval of to-be-remembered material. However, the evidence does not appear strongly to support either hypothesis. The next logical step then in isolating these effects was, as mentioned above, to run a conceptual replication which more or less directly paralleled the methodology employed by the previous researchers in the area, but using in one condition a more balanced forced-choice recognition task than that used by Snyder and Uranowitz, and employing the McCloskey and Zaragoza "modified procedure" in an attempt to

eliminate the impact of response-bias effects. Such an experiment is reported in the following chapter.

CHAPTER 8 - HOW THE RECONSTRUCTION OF THINGS LASTS: DELAY, AND THE EFFECTS OF STEREOTYPIC INFORMATION PRESENTED AT RETRIEVAL

> "At any rate he was certain, if it should prove the fact that such a history was really extant, being that of a knight-errant, it could not be otherwise than lofty, illustrious, magnificent, and true. This thought afforded him some comfort, but he lost it again on considering that the author was a Moor...and that no truth could be expected from Moors, who are all imposters, liars, and visionaries."
> Miguel Cervantes: *Don Quixote de la Mancha*

Experiment 5

Introduction

Experiments 4a and b were attempts experimentally to characterise the effects of stereotypic information presented at retrieval. The results lend some support to the hypothesis proposed by Snyder and Uranowitz (1978c) that such information can act as retrieval schemata for stereotype-consistent information. However, the results are inconclusive, and Bellezza and Bower's rival formulation - that such information induces response-bias in the subjects receiving it - remains a possibility. This issue has great relevance to the investigation of stereotypes and eyewitness memory, since, as outlined in Chapter 7, both formulations imply far-reaching effects. Hence it is important here to specify the role played by stereotypic information in this situation.

Experiment 4a employed a recognition procedure designed to eliminate reponse-bias effects. This procedure was applied to an experimental situation not fully parallel to that used in the original experiments by Snyder and Uranowitz, Clark and Woll,

or Bellezza and Bower. It was argued that the non-employment of a delay between receipt and testing of target material was a fairly conservative departure from the original design. However, it is possible that it might lead to subtle changes in the memory trace (see, e.g., Srull, 1984). Additionally, testing subjects after a delay has more external validity than testing them immediately: rarely will a police interview take place within five minutes of a crime being witnessed. Therefore, a more direct replication of the original studies - still employing the "modified procedure" along the lines of McCloskey and Zaragoza (1985) - is employed in the present experiment. This is a relatively small experiment, the purpose of which is to investigate the effect of delay in order to examine whether this variable might have implications for the results achieved in the various Betty K experiments reported in the literature. This should help highlight the relative efficacy of the response-bias and the retrieval schema hypotheses in explaining the results obtained.

Methodology

Subjects: Eighty subjects, approximately half of them male and half female, aged 18 - 21 years, participated in this experiment. All were undergraduates enrolled in the First Year Psychology course at the University of St Andrews. Allocation to the four experimental conditions (see below) was

carried out on a random basis, with the constraint that there should be twenty subjects per condition.

Materials: These were identical with those used in Experiments 4a and b (see Appendix 18a).

Procedure: Subjects were individually presented with the Betty K biographies and stereotyping paragraphs. Five minutes was allowed for reading this material. Subjects were then presented with the four short-answer questions designed by Bellezza and Bower (1981) to increase the salience of stereotype label (see Appendix 18a).

Subjects in the two immediate testing conditions completed modified procedure recognition tasks immediately after receipt of the four short-answer questions. Subjects in the two delay conditions were tested three weeks after receipt of the experimental materials, (this being the delay employed by Snyder and Uranowitz, 1978c,) again with the modified procedure.

Results

Analysis of variance performed on the entire data set revealed significant differences between the four experimental groups: $F(3,76) = 8.1220$ $p = 0.0001$ (see Appendix 24, Table 1). A significant interaction was also revealed between experimental group and type of information generated in the recognition task: $F(9,228) = 13.9513$ $p = 0.0000$. Finally, a significant

difference was found between the four dependent measures - $F(3,228) = 201.6994$ $p = 0.0000$ - although this was unsurprising, since a glance at, e.g., Figure 8.1 compared with Figure 8.3, will show that many more hits than false alarms were generated.

Two smaller analyses of variance were then performed in an attempt to isolate the cause of these effects.

A comparison of the results achieved under the two stereotype labels revealed a significant difference: $F(1,78) = 5.1058$ $p = 0.0266$. There was also a significant interaction between label received and information generated: $F(3,234) = 2.8173$ $p = 0.0398$. A significant difference was also discovered between the types of information generated - $F(3,234) = 138.3138$ $p = 0.0000$ - but again this was a somewhat banal finding, since, as mentioned above, many more hits were generated than false alarms (see Appendix 24, Table 2). Where subjects were presented with a heterosexual label, they were found to make significantly more heterosexual hits than lesbian hits: $F(1,39) = 49.8400$ $p = 0.0000$ (see Figure 8.1, and Appendix 24, Table 6). A similar effect was found where subjects were presented with a lesbian label for Betty K: surprisingly, a greater number of heterosexual than lesbian hits was also found under this condition: $F(1,39) = 30.8900$ $p = 0.0000$ (see Figure 8.2 and see Appendix 24, Table 7). This result runs counter to the predictions of both rival theories.

FIGURE 8.1 - LESBIAN vs HET. HITS
TWO TEST TIMES, TWO LABELS

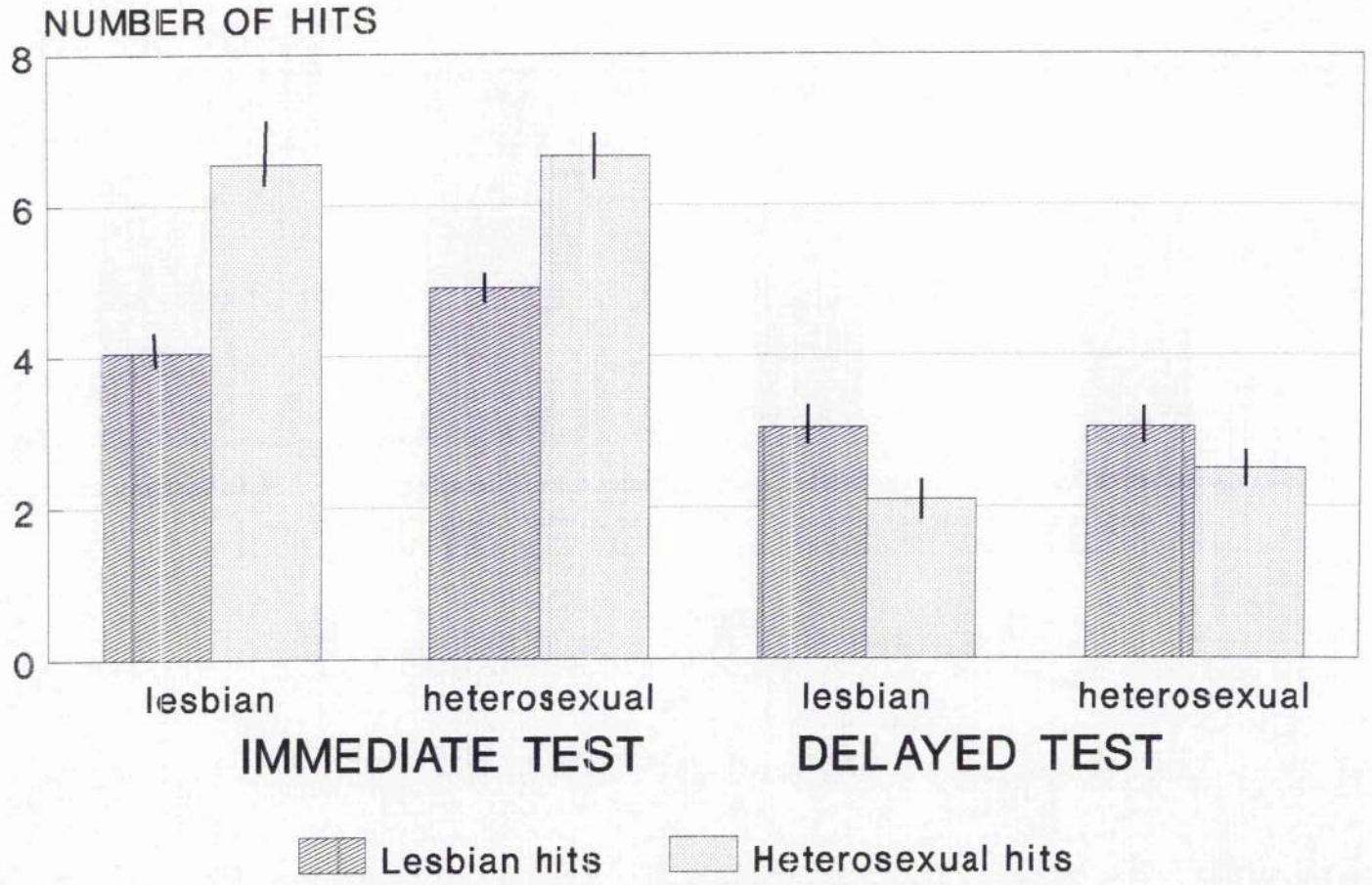
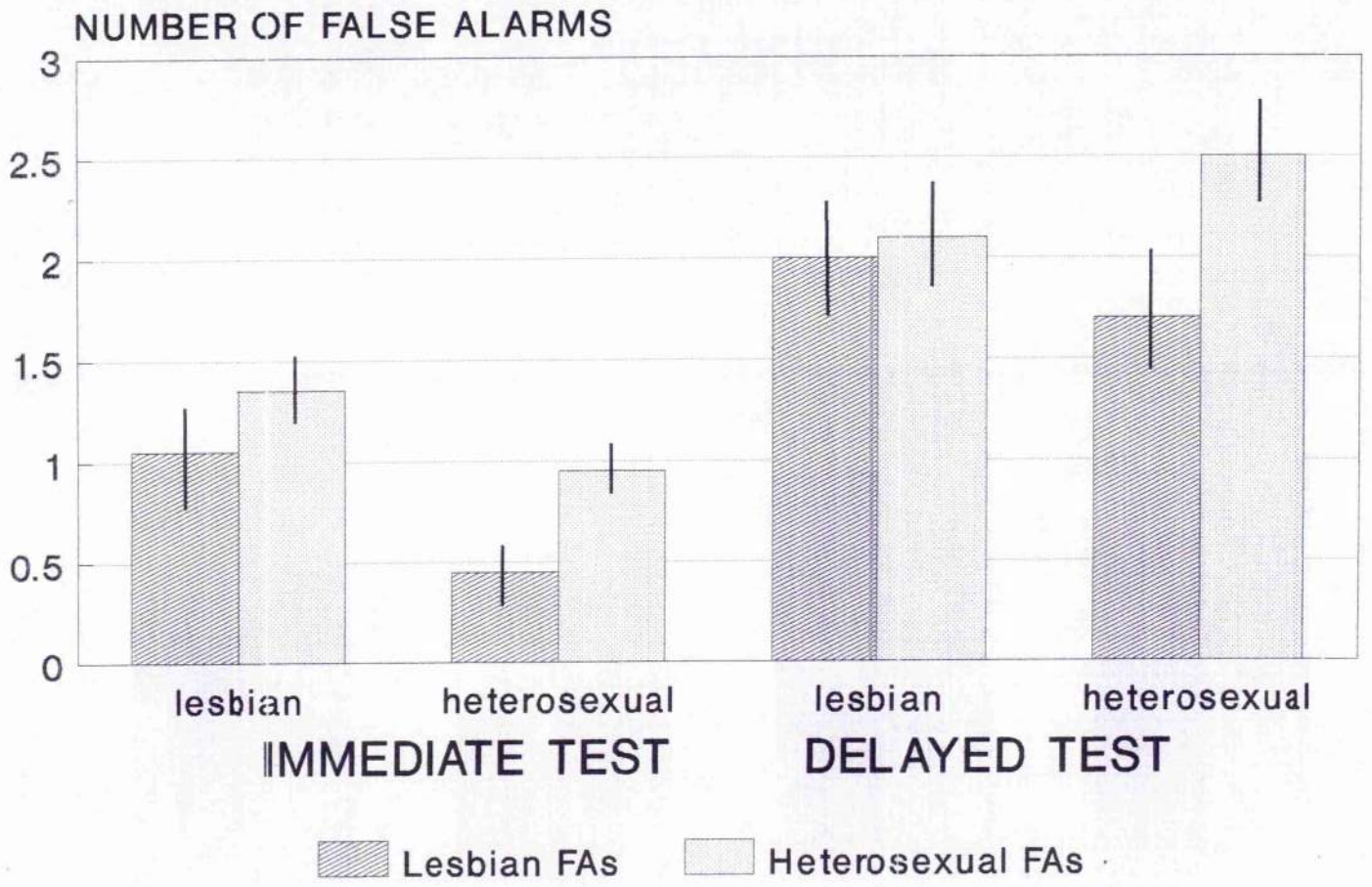


FIGURE 8.2 - LESBIAN vs HET. FAs
TWO TEST TIMES, TWO LABELS



No significant difference was found in the number of false alarms generated under the two label conditions, however: $F(1,78) = 1.0181$ $p = 0.3161$, although over all a greater number of heterosexual than lesbian false alarms was made: $F(1,78) = 1.6037$ $p = 0.2910$. (This data is presented in Figures 8.3 and 8.4, and see Appendix 24, Table 8). This finding lends more support to the retrieval-schema theory than to the response-bias theory, since the latter predicts that a greater number of false alarms will be made when a lesbian label is presented than when a heterosexual label is presented.

A comparison of the results achieved under the two delay conditions revealed a significant difference: $F(1,78) = 9.8831$ $p = 0.0024$; and also a significant interaction between delay condition and information generated: $F(3,234) = 32.3037$ $p = 0.0000$. Again, a significant difference was discovered between the types of information generated: $F(3,234) = 188.7778$ $p = 0.0000$ (see Appendix 24, Table 3).

A Neumann-Keuls test indicated that there were significantly more heterosexual hits made under a heterosexual label than lesbian hits made under a lesbian label ($Q = 1.458$ $p < 0.05$). Again, this is a surprising result which would not be predicted by either theory.

Similarly, a series of Neumann-Keuls test was performed upon the results for the comparison of delay conditions.

FIGURE 8.3 - NEUTRAL HITS, TWO TEST TIMES, TWO LABELS

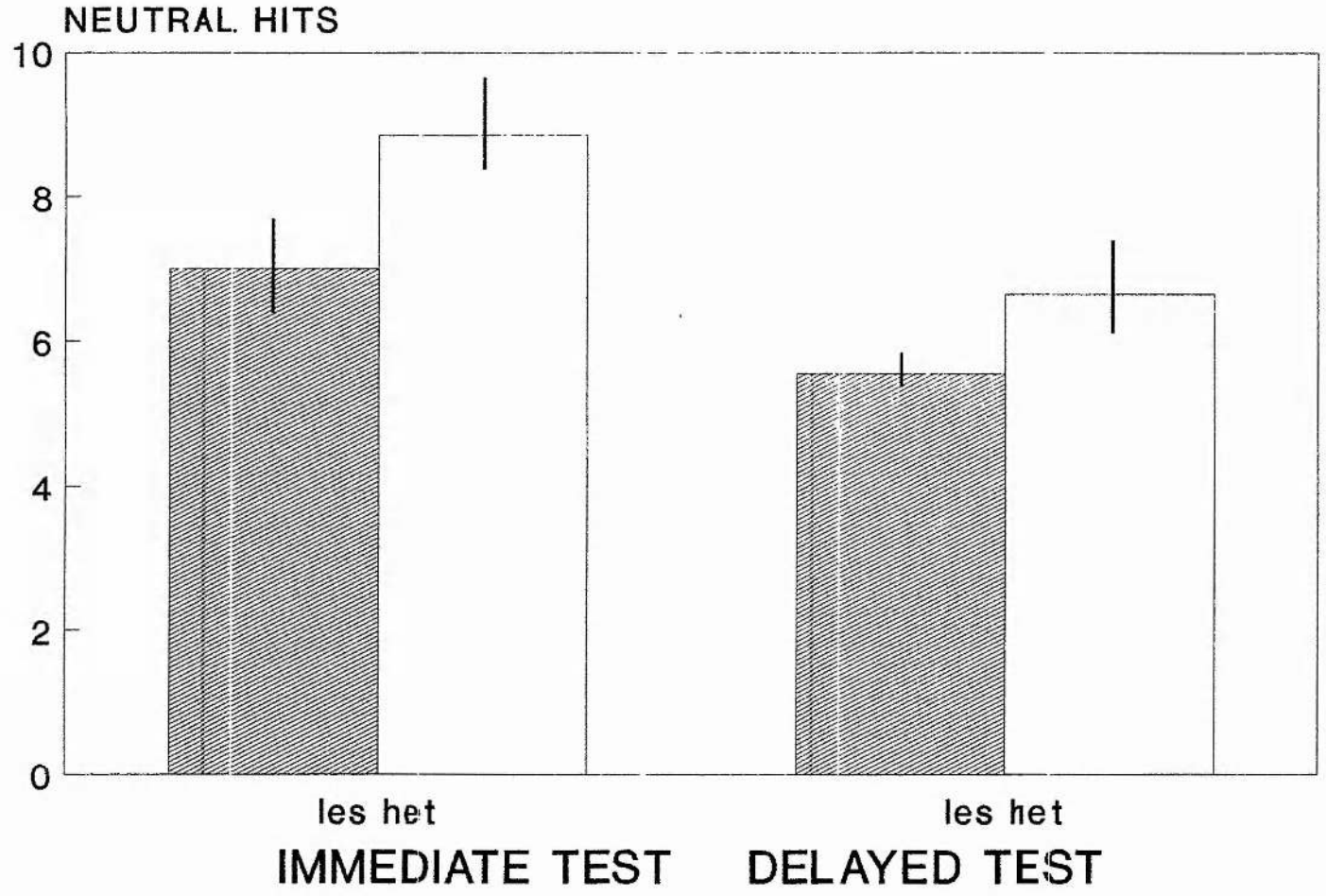
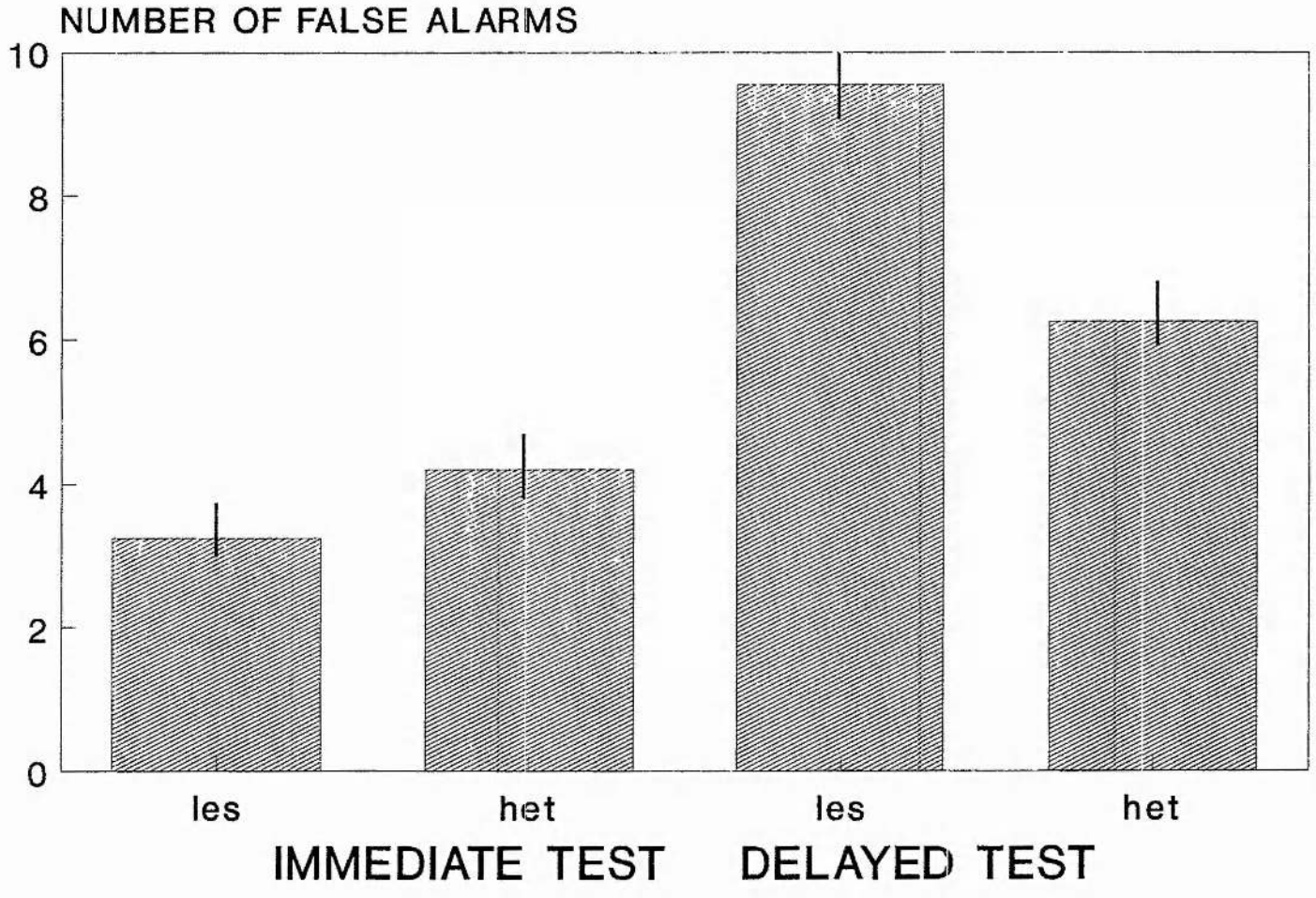


FIGURE 8.4 - NEUTRAL FALSE ALARMS, TWO TEST TIMES, TWO LABELS



The results were however uninteresting: the only significant differences found were between numbers of hits and false alarms generated.

Analyses of variance were also performed on the number of neutral hits and false alarms generated under the two label conditions and the two delay conditions.

There was a significant difference for neutral information generated across the two label conditions: $F(1,78) = 5.6945$ $p = 0.0194$; however stereotype label appears to have had no independent effect upon neutral hits or false alarms: $F(1,78) = 1.0458$ $p = 0.3096$. Investigating the effects of delay, there was a significant main effect of delay: $F(1,78) = 5.9605$ $p = 0.0169$, and also a significant interaction between delay condition and neutral information generated: $F(1,78) = 56.0119$ $p = 0.0000$. Significantly more neutral hits were generated in the immediate testing condition than in the delayed testing condition; while significantly more neutral false alarms were generated in the delayed than in the immediate testing condition, (see Appendix 24, Tables 4 and 5).

Discussion

The present experiment was run in an attempt to ascertain whether the differences between Experiments 4a and b and the results of previous researchers, Snyder and Uranowitz (1978c), Bellezza and Bower (1981), and Clark and Woll (1981), were due to the fact that these researchers included in their

experiments a delay between receipt of stereotypic information and the testing of recognition memory. Although a main effect of delay was discovered in the present experiment, this was not found to have a significant effect upon recognition memory for any of the measures that were of particular interest in Experiment 4. This parallels the finding of Lutz (1983), who found evidence of stereotypic bias due to a label presented at retrieval one week after encoding of target material.

The results of Experiments 4a and b tended to support the retrieval schema theory rather than the response-bias theory. The present experiment indicates that the differences between the results of Experiments 4a and b and those of other researchers who have conducted Betty K-type studies was not due to the inclusion in previous studies of a delay between receipt of target material and testing. Therefore, as far as this major aim of the experiment is concerned, the findings of Experiment 4 are supported: the retrieval-schema theory continues to appear a superior explanation of the results than the response-bias theory.

Particularly damaging for the retrieval-schema theory in Experiments 4a and b was the finding that presentation of a lesbian stereotype label failed to improve subjects' recognition memory for lesbian stereotype-consistent material. This finding does cast doubt over the validity of the theory. Other predictions of the theory did however receive support.

The rival formulation, proposed by Bellezza and Bower, was not supported in Experiments 4a and b. It was a surprising result of the present experiment that, under lesbian-label conditions, subjects generated significantly more heterosexual hits than lesbian hits. This finding runs directly counter to the retrieval-schema theory. Additionally, where a heterosexual label was given, more heterosexual than lesbian hits were generated, where neither theory would predict a difference. These results therefore count against both theories. Indeed, the former result seems to imply the existence of an *in*consistency effect of stereotypic information presented after target material. This is in direct opposition to the predictions of Snyder and Uranowitz and Bellezza and Bower, and also those of Wyer, Bodenhausen and Srull (1984), and Stangor and Ruble (1989). These latter two papers, it has been argued, present probably the most convincing models of the effects of stereotypic information presented at retrieval.

An amendment to the retrieval-schema theory might make its predictions consistent with the results achieved here, however: The discovery that more heterosexual hits are generated under a heterosexual label is consistent with the retrieval-schema theory if it is held that "heterosexuality" could act as a stereotype label. However, it has been argued that Snyder and Uranowitz' results appear to make this unlikely. Snyder and Uranowitz found no difference in

the pattern of results given by subjects in a heterosexual-label condition and those given by subjects in a no-label condition, implying that to say that a target was "heterosexual" was tantamount to providing no stereotypic information about the target.

In attempting to explain these results it is important to note that, across conditions, more heterosexual than lesbian false alarms were made. It might be argued by a proponent of the retrieval-schema theory that this represents the schema action of the "lesbian stereotype" in reducing the number of "lesbian false alarms" made in this condition when this stereotype was presented. Alternatively, it could be argued that, when unable to remember a correct bit of information, subjects may have relied upon conventional societal norms to provide them with their answer and, hence, were more likely to choose an item consistent with a representation of Betty as a heterosexual than as a homosexual. Intuitively appealing though such an explanation may be, however, it is probably not tenable. The explanation is a virtual reversal of the argument presented by Bellezza and Bower (1981) in their critique of the Snyder and Uranowitz experiment; and the forced-choice recognition task that was employed in the present experiment was designed with the intention of eliminating response-bias.

Discounting the present experiments, the literature contains one successful replication (Lutz, 1983) and two unsuccessful replications (Bellezza and Bower,

1981, and Clark and Woll, 1981) of the Snyder and Uranowitz study. The literature also boasts at least two theories of social memory which predict a consistency effect of stereotypes presented at retrieval; an effect of the type discovered by Snyder and Uranowitz. It would certainly be premature to dismiss, on the basis of the present study alone, the theory that stereotypic information might act as retrieval schemata. Wyer *et al* (1984) found a consistency effect in this paradigm; as did Lutz (1983) and Snyder and Uranowitz (1978 a, b, and c). Bellezza and Bower (1981) also found some effect, although they attributed this to response-bias. Stangor and Ruble (1989) present a model of social memory which implies a consistency effect in this paradigm. Experiments 4a and b in the present thesis found evidence to support the theory that stereotypic information presented at retrieval will act as retrieval schemata for stereotype-consistent information. Against this evidence, Experiment 5 appears to indicate that an inconsistency effect may be more likely: although results are also reported which appear to add weight to the retrieval-schema theory. The only studies which fail to find any effect when stereotypic information is presented after target information are those by Rothbart, Evans and Fulero (1979) and Clark and Woll (1981). The most conservative appraisal of the state of the literature might be to say that it does appear to show that stereotypic information guides or distorts retrieval

of material from social memory when presented at retrieval of that material. From the limited number of studies extant, however, it is difficult to make powerful predictions about the action of stereotypic information presented at retrieval in different situations, although it appears more likely to find a consistency effect than an inconsistency effect in most studies.

Conceivably, however, the exact type of result achieved might be found to depend, to some extent at least, upon the particular stereotype and encoding set employed, and so on. Srull (1984) argues a parallel case for the effect of stereotypic information presented at encoding, in which he claims that different experimental conditions are likely to produce different effects. This is certainly an area which future work should consider. It is particularly important to the present thesis to test the effect of stereotypic information presented at the retrieval in a situation which parallels as closely as possible a real-life crime situation. This is the area which will be addressed in the next chapter.

CHAPTER 9 - KISS AND TELL: THE EFFECT OF STEREOTYPIC INFORMATION ON MEMORY FOR A STAGED INCIDENT

"You must remember this:
A kiss is just a kiss"
Max Steiner: As Time Goes By

Introduction

The results of Experiments 4 and 5 appear to indicate that it is unlikely that stereotypic information could be used in the police interview situation as a means of increasing the amount of correct information generated by an eyewitness. The present chapter continues to investigate the effects of stereotypes presented at the retrieval of target material. An experiment is reported in which stereotypic information is presented during the retrieval of information from an incident analagous to a crime in that it is dramatic and unexpected. Additionally, the effects of very long (six-month to one year) retention intervals are considered.

All of the experiments reported so far have been laboratory-based, i.e., both presentation and testing occurred in a well-regulated, artificial setting. Perhaps more importantly, target materials were somewhat artificial. While it is an important aspect of eyewitness research to generate and refine hypotheses in the laboratory, it is also necessary to test these hypotheses using situations as close as possible to genuine crime situations.

For obvious ethical reasons, effects isolated in the laboratory can rarely be directly tested in crime situations. However, it is possible to utilise a

paradigm that falls roughly half-way between the artificiality of the laboratory study and the realism of the field study: this is the "staged incident". The experiment reported here, then, attempts to broaden the range of research that has been conducted into the effects of stereotypic information on memory. It attempts to specify more directly what type of effect might be expected to hold when stereotypic information is presented during retrieval of information about a criminal episode: specifically, during a police interview.

An Outline of Experiment 6

The "staged incident" ("*Wirklichkeitsversuch*", "reality experiment", or "event test") has had a long and distinguished history in the service of investigators into eyewitness memory. As Clifford and Bull (1978) remark:

> "This experimental approach is a big advance over static inspection and recognition of photographs in the sense that the action is dynamic, it reproduces real life in content and subjects' unpreparedness for becoming witnesses, and, because it is experimental, very careful checks can be made on all aspects of perception, memory and identification processes especially the rating data given by witnesses as to their certainty of correctness." (Clifford and Bull, 1978, p 52.)

In this particular instance, the authors are especially concerned with memory for faces. However, the general thrust of their remark holds true for all aspects of eyewitness memory.

The aim of this experiment is to test the effects of stereotypic information on eyewitness memory for an

incident with greater external validity than the materials employed so far.

For the purposes of this experiment, the usual form of staged incident, (in which, e.g., a rack of technical intruments is maliciously knocked over - Malpass and Devine, 1984a - or a calculator is surreptitiously slipped into a thief's pocket - Wells, Leippe and Ostrom, 1979,) was considered inappropriate. It was necessary to design an incident with a number of special characteristics. The staged incident is generally unexpected, long-lasting, and capable of being performed in front of a large number of witnesses. This particular incident, however, also had to be quite dramatic (so as to resemble as far as possible an actual crime situation) and susceptible to the influence of stereotypic information.

To these ends, it was decided to interrupt a (well-attended) first-year social psychology lecture with the appearance of a kissogram; following this incident with a verbal altercation between the kissogram and her disapproving "boyfriend". The incident contained a number of characteristics judged to be "typical" or "untypical" of kissogram incidents (see below). Also included was a number of characteristics judged to be "neutral" with respect to kissogram incidents. Interviews with witnesses ascertained the effect of stereotypic information on memory for these types of information.

Experiment 6 considers two other aspects of the eyewitness memory situation which, while potentially

important, have received little attention in the literature. These are witnesses' memory for clothing, and the decay of the eyewitnesses' memory trace with time. A brief review of each of these areas is given below.

MEMORY FOR CLOTHING Clothing is potentially an extremely important feature of the crime situation. Not only is it a salient contextual cue to recall (Cutler *et al*, 1986) - and hence an important variable to control in the identification parade - it might also be used to provide leads in criminal investigations. Very few studies have investigated memory for clothing however. In a review of the literature, MacLeod and Frowley (1991) were able to find just three studies in which a target person's clothing had been altered for recognition testing; and none in which a recall measure had been used. The only exception to this generalisation is the work of Yuille and his colleagues, which is addressed below.

The studies in which recognition memory was tested were all performed by Thomson and his colleagues, (Thomson and Robertson, 1980; Thomson, 1981; Thomson, Robertson and Vogt, 1982, - reviewed in chapter 2, above). Thomson found that keeping the variables (as he called it, the "context" - which included the clothing variable) constant had a massive effect on recognition accuracy. In Thomson (1981), for example, accuracy increased from 25% to 89%.

These studies indicate the potential importance to eyewitness memory of such contextual variables as

clothing; although they failed to show that clothing *per se* has a great effect. Thomson *et al* (1982) found a decrement in recognition performance when all three context variables (i.e., clothing, activity performed, and physical setting) were altered (Experiments 1 and 2), but that this was not attributable to an independent change in any one variable.

MacLeod and Frowley (1991, p 10) conclude that: "These studies indicate that covarying context variables may have some enhancing effect on recognition accuracy. Whether any single context variable would have such an effect was not investigated but cannot be ruled out." Since clothing was deemed by Thomson to be a context cue, MacLeod and Frowley felt it appropriate to present an account of the important work on context cues performed by Cutler, Penrod and their associates (Cutler and Penrod, 1988; Cutler, Penrod and Martens, 1987a and b; Cutler, Penrod, O'Rourke and Martens, 1986), in an effort to determine the importance of context cues in eyewitness situations.

Essentially, these studies determined that context cues are effective only when manipulated in multiple fashion, i.e., when more than one cue is manipulated at a time. The implication, therefore, is that a change in the clothing variable alone may have a very limited effect on eyewitness performance.

The other central consideration is this: Just how good is memory for this particular context cue?

Again, the psychological literature has difficulty in answering this question. The only relevant studies are those by John Yuille and his colleagues (Yuille and Cutshall, 1986/1989; Tollestrup, Turtle and Yuille, in press). These studies considered witnesses' memories for real-life crimes, with the clothing variable considered only incidentally. Yuille and Cutshall found that:

> "...colors were correctly provided 66% of the time in the police interview [ie. the original interviews administered by police after the crime in question had taken place] and 59% to the researchers, while style was correct 88% and 80% respectively. Thus, the color of clothing seems to be the most difficult feature to retain (or notice). For example, one witness graphically described the wounds on the body of the thief and she provided detailed descriptions of the body's position and its exact location in the street. Whereas this information was highly accurate, she erroneously described him as wearing a T-shirt and a red and black plaid jacket. He actually wore a dark blue sweater and a blue jean jacket." (Yuille and Cutshall, 1986, p 296.)

This paper provides the only systematic study so far of memory for clothing. Yuille and Cutshall's results provide two hypotheses for the present experiment: first, that memory for *style* of clothing should be superior to memory for *colour*; second, that memory for both style and colour should fade with time. It should be mentioned however that some doubt may be cast over Yuille and Cutshall's results, given that the only subjects used in their experiment were volunteers - who may therefore have had greater confidence in their ability to remember information from the witnessed crime than does the "average" eyewitness. While the points established by Yuille and Cutshall regarding clothing may be importantones

for the police investigator to be aware of when deciding how much credence to give an eyewitness' account of the style and colour of a suspect's clothing, they are not of central importance to the present study.

More important is a third hypothesis, derived from the social memory literature: that memory for the kissogram girl's clothing should be enhanced where stereotypic information is presented; memory for her "boyfriend"'s clothing however - being incidental information - should remain unaffected. This is the most important of the hypotheses concerning memory for clothing.

THE EFFECT OF TIME In a genuine eyewitnessing situation, a witness may be interviewed many times over the course of several months (Roy, 1991). During this time, the memory trace may degrade, although some questions about this remain unanswered, e.g., what is the relative rate of decay of different aspects of the memory trace, for how long is the memory trace reliable, and so on. Conceivably, the situation may be even more complicated that this. Witnessing a crime may cause the spontaneous generation of what Brown and Kulik (1977) term "flashbulb memories". Although the accuracy of such memories has not been tested, subjectively they appear highly detailed and remarkably stable over time. This implies that memory for crimes may decay at a much slower rate than memories for more mundane occurrences. Even if a

flashbulb memory is not formed, it is possible that the repetition of the story of the crime (with friends, relatives and so on) over time may serve to keep the incident fresh in the mind. Alternatively, embellishment of details may take place during story-telling, thus causing erroneous information to be kept "fresh" in the mind.

There is a certain controversy in the eyewitness memory literature as to the likelihood of these different effects. The most common opinion is the intuitive one that the memory trace degrades with time (see, e.g., Penrod, Loftus and Winkler, 1982); however, Yuille and Cutshall (1986) take issue with this claim (although one should bear in mind that their subjects were volunteers from the set of witnesses to the crime; and those with poorer memories may possibly have "deselected" themselves):

> "One of the more striking results was the lack of memory loss over time. Accuracy rates remained virtually unchanged five months after the incident. This results calls into question the general conclusion that eyewitness memory deteriorates rapidly with time...an Ebbinghaus decay curve simply doesn't apply in this type of case."
> (Yuille and Cutshall 1986, p 299.)

Only a few studies in the eyewitness literature explicitly consider this variable. Deffenbacher and Horney (1981) conclude that, although there are methodological problems with much of the earlier work, a reasonable estimate of average accuracy loss would be about 0.3% per day. Estimates vary somewhat about this figure (cf. Marshall, 1966; Lipton, 1977); however, Deffenbacher and Horney agree with Yuille and

Cutshall that the forgetting curves are less steeply declining than one might predict from Ebbinghaus forgetting functions.

Hypotheses: To summarise, the hypotheses for the present experiment are:

1. presentation of stereotypic information will aid memory for stereotype-consistent information, but not for other forms of information. A corrollary of this is that presentation of stereotypic information will aid memory for the kissogram's clothes but not the "boyfriend"'s. Additionally, memory for style of clothing should prove superior to memory for colour of clothing. Memory for both style and colour will deteriorate with time.
2. there will be *less* deterioration of memory with time in the stereotype condition than the non-stereotype condition for stereotype-consistent information
3. the total amount of information recalled will decrease with time. The accuracy of memory will also decrease with time. Accuracy will be greater for stereotype-consistent than -inconsistent information in the stereotype condition.

Methodology

Generation of Materials: Fifty first-year Psychology students were asked to list events or characteristics which they considered highly typical or highly untypical of a "kissogram artiste" and kissogram

incident. They were also asked to write down the noun which they considered most appropriate for the stereotyping of a "kissogram artiste". Thirty of fifty subjects suggested the word "bimbo". This 60% level of agreement between subjects is higher than the 25% agreement suggested by Bower, Black and Turner (1979) as constituting a shared stereotype, or even the 50% value, which Mortimer (1991) suggests may be more meaningful. No other noun was suggested more than three times. The data thus generated was used to provide an outline for the staged incident. The four most stereotype-consistent and -inconsistent characteristics were both included as items of to-be-remembered material e.g., the incident was accompanied by music, since this was considered typical of a kissogram incident; but the kissogram artiste was wearing glasses, since this was considered untypical. Also included were four items of incidental material, which were neutral with respect to the kissogram stereotype.

Two actors (one male, semi-professional; one female, amateur) were employed to stage the incident. Several rehearsals took place in the appropriate lecture theatre before the incident proper took place (for a description, see Appendix 25).

Subjects: The incident was staged during a lecture attended by approximately one-hundred and fifty first-year undergraduate Psychology students. Twenty-five subjects volunteered to take part in the first set of

interviews, which occurred one to two days after the incident; sixteen of these took part in the second set, six months after the incident; and twelve in the third set, one year after the incident. Of the twenty-five subjects who took part in the first set of interviews, seventeen were female and eight male. All subjects were aged between seventeen and nineteen years at the beginning of the study.

Design: This was a between-subjects design: half of the subjects were presented with stereotypic information (see below) during the retrieval of information about the staged incident; half were not.

Procedure: On the day following the staged incident, an announcement was made to the social psychology class that the event of the previous day had been staged, and that it would be appreciated if people would attend the Social Psychology Laboratory the next day to be interviewed about their memory of the incident. Each subject was met in the foyer of the department and escorted to the laboratory. The interviews took place at fifteen minute intervals, and took about ten minutes each to complete.

During the interview, each subject was simply asked to recall as many details as possible about the incident. They were asked to report details in "roughly" chronological order, and not to worry about how important or trivial the reported details might appear. Occasionally, when a subject skipped over a

particular detail, or omitted it entirely, then the interviewer would ask a direct question concerning that detail, as in a standard police interview (Roy, 1991). Otherwise the interviewer remained silent.

Once the subject had finished recounting the incident to his/her own satisfaction, s/he was asked to "think through" the incident once more, and report any new details that s/he remembered and had not reported during the first stage of the interview.

Approximately six months later, subjects returned for a second interview. They were met in the foyer by an accomplice who escorted them to the laboratory. On the way, the accomplice feigned ignorance of the experimental design, asking the subject, "I don't know anything about this experiment, is it the one where the girl came into the lecture hall and took her clothes off?" - substituting, in the stereotype condition, the word "bimbo" for "girl". In this way, it was hoped to reproduce the effect of a stereotype being dropped during the course of a police interview, but without the experimenter being aware of the group to which any given subject belonged. Once the subject had entered the laboratory, the interview structure was identical with that used previously.

The third interview, (again taking place after a six-month interval,) followed the same format as the second, except that this time the accomplice dropped no stereotype-related information.

All interviews were tape-recorded for later analysis.

Analysis: Due to the substantial subject drop-out rate in each condition, the data were analysed using unequal-n analysis of variance. Three categories of information were of particular interest. First were neutral, incidental items, e.g., the subject that the girl was supposedly studying. Second were four stereotype-consistent items, e.g., removing the garter and handing it to the "victim". Third were four stereotype-inconsistent items, e.g., wearing glasses.

Hence, during analysis, it was possible not only to compare total numbers of items generated under each condition, but also the type of items generated across conditions, e.g., were more stereotype-consistent items recalled in the stereotype condition than in the non-stereotype condition?, and also to compare the types of item recalled within conditions, e.g., were more stereotype-consistent than -inconsistent items recalled in the stereotype condition?

Results

The raw data for correct information generated are represented in Table 9.1.

Hence, total level of recall did drop, as predicted, across the one year study period: from 10.857 to 10.0 in the stereotype condition, and from 11.4 to 9.7 in the non-stereotype condition. Contrary to expectations, the decline does not appear to have been steady. Total recall levels at the second interview

Delay		Group 1	Group 2
1	M	10.857	11.400
	SD	3.388	2.989
2	M	9.429	8.500
	SD	7.032	4.389
3	M	10.000	9.700
	SD	1.732	1.252

TABLE 9.1 - MEAN AMOUNT OF CORRECT INFORMATION RECALLED x INTERVIEW GROUP x DELAY INTERVAL

are in fact slightly *lower* than those at the third, although this difference is not significant: $t(6) = -0.56$. Analysis of variance showed that this overall decrease in recall level with time was indeed significant: $F(2,30) = 3.4759$ $p = 0.0439$ (see Appendix 26, Table 1). However, delay did not have a significant effect on memory for any of the three types of information particularly tested: $F(2,30) = 2.3791$ $p = 0.1099$ for stereotype-consistent information; $F(2,30) = 0.8071$ $p = 0.4556$ for stereotype-inconsistent information; $F(2,30) = 1.0012$ $p = 0.3794$ for incidental information (see Appendix 26, Tables 2 - 4). No significant effect was found for either group alone: For Group 1, $F(2,10) = 1.4286$, $p = 0.2846$; for Group 2, $F(2,12) = 1.2000$, $p = 0.3349$ (see Appendix 26, Tables 5 and 6). Hence, the cumulative effects of a number of small, non-significant decrements in memory was sufficient to produce a significant overall decrease in the number of items recalled.

The effect of presentation of stereotypic information on memory was also tested. It was not found to affect total level of recall: $F(1,15) = 0.0825$ $p = 0.7778$. Neither was it found to affect recall of stereotype-consistent information, $F(1,15) = 2.7574$ $p = 0.1176$, stereotype-inconsistent information, $F(1,15) = 0.846$ $p = 0.3722$, or incidental information, $F(1,15) = 0.9148$ $p = 0.3540$ (Appendix 26, Tables 1 - 4).

Finally, interactions between delay and presentation of stereotypic information were considered for each type of information; none was significant. $F(2,30) = 0.4001$ $p = 0.6738$ for total information recalled; $F(2,30) = 2.8793$ $p = 0.0718$ for stereotype-consistent information; $F(2,30) = 0.1566$ $p = 0.8558$ for stereotype-inconsistent information; $F(2,30) = 0.3612$ $p = 0.6998$ for incidental information (Appendix 26, Tables 1 - 4).

Next, analysis of variance was employed to investigate within-group effects. For subjects receiving stereotypic information, more stereotype-consistent information was recalled than stereotype-inconsistent information: $F(1,18) = 64.8$ $p = 0.0000$. There was no significant interaction between delay and consistent/inconsistent information recalled: $F(2,18) = 0.6000$ $p = 0.5594$ (Appendix 27, Table 1). The same pattern of results held for non-stereotype subjects. More stereotype-consistent information was recalled than stereotype-inconsistent information: $F(1,18) = 10.2857$ $p = 0.0034$, and there was no interaction

between stage of interview and information recalled: $F(2,27) = 2.7857$ $p = 0.0795$ (Appendix 27, Table 2).

Hence presentation of stereotypic information appears not to have affected the amount of correct information generated. The next question to be considered was the number of errors made. The raw data are presented in Table 9.2.

Analysis of variance was performed on these data. Delay did not affect the number of errors made in the recall of stereotypically -linked items (i.e., stereotype-consistent or -inconsistent items): $F(2,30) = 0.9696$ $p = 0.3908$; $F(2,30) = 3.0612$ $p = 0.1006$ respectively.

Delay		Group 1	Group 2
1	M	1.714	2.300
	SD	1.704	0.949
2	M	1.429	2.600
	SD	0.976	4.761
3	M	0.857	1.100
	SD	1.215	0.9944

TABLE 9.2 - MEAN NUMBER OF ERRORS MADE x INTERVIEW GROUP x DELAY

Delay was however found to affect the number of errors made in recall of incidental items: $F(2,30) = 3.9819$ $p = 0.0293$. There was no interaction between time and presentation/non-presentation of stereotypic information for any type of information: $F(2,30) = 0.0302$ $p = 0.9703$ for stereotype-consistent information, $F(2,30) = 1.6100$ $p = 0.2167$ for

stereotype-inconsistent information, $F(2,30) = 0.6180$ $p = 0.5457$ for incidental information (see Appendix 28, Tables 1 - 3).

Accuracy: Next, the accuracy (i.e., pieces of correct information ÷ total pieces of information) of recall in the two stereotype conditions was calculated. Only *total* information and *incidental* information were considered. These data are presented in Table 9.3. Stereotype-consistent information was not considered because no errors were made by any subject here (hence the total scores presented above may be considered as accuracy scores). Stereotype-inconsistent information could not be considered because many scores were not amenable to statistical analysis. In some cases, a number of errors was made, although no items of correct information were produced.

Delay		Group 1	Group 2
1	M	87.510	83.123
	SD	40.489	5.914
2	M	87.159	77.369
	SD	7.810	12.847
3	M	93.333	90.967
	SD	8.598	10.529

TABLE 9.3 - ACCURACY OF INFORMATION RECALLED x INTERVIEW GROUP x DELAY CONDITION

There was a significant main effect of time on accuracy of recall of total information: $F(2,45) =$

4.2754 p = 0.0200. A Neumann-Keuls test applied to these data indicated that accuracy was greater after a delay of one year (91.951%) than after six months (81.400%, Q = 10.5568, p < 0.01) or during immediate testing (84.9297%, Q = 7.0212, p < 0.05). Although there was no main effect of delay on the number of errors made in any category of information, this effect does appear likely to be due to an overall decrease in errors made with delay time. The mean number of errors made fell from 0.121 to 0.000 for stereotype-consistent information; from 0.314 to 0.286 for stereotype-inconsistent information; and from 2.007 to 0.979 for incidental information. There was also a slight concomitant fall in the number of items correctly recalled in each category with delay. Presentation of a stereotype was not foundto affect overall accuracy (i.e., across *all* information types), although this main effect only missed significance by seventy-four thousandths: F(1,45) = 3.8040 p = 0.0574. The trend was towards greater accuracy in the stereotype than in the non-stereotype condition (89.340% in the stereotype condition against 83.820% in the non-stereotype condition). There was no significant interaction between the two: F(2,45) = 0.6106 p = 0.5475. There was no main effect of stereotype presentation on accuracy of recall of incidental information: F(1,20) = 1.0679 p = 0.3138; no main effect of time: F(2,20) = 0.7515, and no significant interaction between the two variables:

$F(2,20) = 0.7515$ $p = 0.4845$ (Appendix 28, Tables 1 and 2).

Therefore, neither delay interval nor presentation of stereotypic information appears to have affected the number of correct or incorrect pieces of information generated, or the accuracy of the recall accounts.

Clothing: The most important hypothesis concerning clothing was that accuracy of memory for the kissogram's clothing would be greater with the presentation of stereotypic information. This was not found to be the case. Accuracy of memory for the kissogram girl's clothing was not affected by presentation of stereotypic information: $F(1,15) = 0.4823$ $p = 0.4980$. Neither was there any interaction between presentation of stereotypic information and delay on memory for clothing: $F(2,50) = 0.7276$ $p = 0.4914$. Accuracy of memory for the "boyfriend"'s clothing was however affected by presentation of stereotypic information: $F(1,15) = 7.1460$ $p = 0.0174$. There was no significant interaction between presentation of stereotypic information and delay on memory for clothing: $F(2,30) = 1.7404$ $p = 0.1927$ (Appendix 30, Tables 1 - 4).

Other hypotheses were advanced concerning memory for clothing. These were less central to the present issue. For the kissogram's clothing, memory for style was more accurate than memory for colour: $F(1,48) = 9.3603$ $p = 0.0036$. Memory for style/colour did not

interact with delay: $F(2,48) = 0.2267$ $p = 0.7980$; neither did delay have any independant effect on the level of memory for style/colour of clothing: $F(2,48) = 0.2401$ $p = 0.7875$.

The same pattern of results held true for the "boyfriend"'s clothing. Memory for style was more accurate than memory for colour: $F(1,48) = 96.0831$ $p = 0.0000$. There was no interaction between memory for style/colour and delay: $F(2,48) = 1.0805$ $p = 0.3475$; neither did delay have any independent effect: $F(2,48) = 2.4849$ $p = 0.0940$.

Discussion

Each of the experimental hypotheses is discussed below.

1) Under no condition was presentation of stereotypic information found to affect level of recall. This is surprising given the number of studies which have found presentation of stereotypic information to affect recall in one way or another (see the meta-analysis table in Chapter 6). Given this fact, perhaps the most valid conclusion that can be reached is that the stereotypic information in the present study was either insufficiently powerful to elicit effects, or that it was not presented in a powerful enough *manner*: Dropping a single word into a (supposedly) chance conversation may simply be an insufficiently forceful medium for the generation of systematic effects. The effects of stereotypic information presented at retrieval are generally less

profound than when presented at encoding. This result is nevertheless a failure to support the theories of Wyer, Bodenhausen and Srull (1984) and Stangor and Ruble (1989): both of which would predict a consistency effect with the methodology employed here.

It was the major purpose of the present experiment to discover the effects of a stereotype "dropped by chance" during retrieval of information about a witnessed crime. It appears unlikely that very large effects will occur under such conditions. However, it should be mentioned that the present experiment used (relatively) low-key stimuli and that the interviews were less of an ordeal for the subjects than the average police interview for a witness. Under more emotive circumstances, somewhat different effects may be found to obtain. Again, though, it appears unlikely that very dramatic effects would be produced.

Memory for clothing was also studied in this experiment. For both targets, style of clothing was found to be more memorable than colour. This accords with the previous findings in the literature (Yuille and Cutshall 1986/89; Tollestrup, Turtle and Yuille, in press) and points to the reliability of this effect. Presentation of stereotypic information was not found to aid memory for the kissogram's clothing. It was however found to aid memory for the "boyfriend"'s clothing. This is the opposite result to that predicted. Since the stereotype presented applied only to the kissogram artiste herself, it is hard to see how this result can be explained. The

best explanation may be no explanation: the result may be due to chance.

2) Under no circumstance was a significant interaction discovered between delay and stereotypic information on information recalled.

3) Although there was a significant effect of delay on total recall level, this effect was not significant in the information categories of specific interest here: i.e., consistent and inconsistent information. The implication therefore appears to be that memory deteriorated slightly, but non-significantly, with time in each category. Hence there appears to be nothing special about stereotypically-linked information: memory for the two categories of stereotypically-linked information (i.e., stereotype-consistent and -inconsistent) appears to have deteriorated no more and no less with time than did memory for other categories of information. It is possible that the subtle effects of stereotypic information were obscured by the other noise present in the results.

The effect of delay on memorial accuracy was found to be completely opposite to that hypothesized. *Total* accuracy was found to increase with time; and this effect appears to have been due to a slight decrease in the number of errors made. It appears that as the number of errors made decreased, then the number of items correctly remembered did not decrease proportionately. Conceivably the false "memory" trace created by erroneous information is less powerful than

that created by correct information, and hence fades more rapidly. Although the subject may indeed frequently have rehearsed the story of the kissogram incident with his/her friends throughout the study period, this rehearsal did not, as might have been expected, lead to greater confabulation; neither does it seem to have been sufficiently powerful to keep the events completely fresh in the mind. Conceivably, delay may cause hypermnesic effects: although the total amount of information recalled appears to decrease with time (across such long delay periods as those used here,) overall accuracy appears to increase. To establish the exact nature of these effects, a separate experiment would be required; and one that is outside the scope of the present thesis. Accuracy was not affected by the presentation of stereotypic information; although a slight trend was discovered in this direction. Had this proven significant, it would have provided support for Snyder and Uranowitz' hypothesis that stereotypic information acts as retrieval schemata. It would not however have supported the contention of Wyer, Bodenhausen and Srull and of Stangor and Ruble that stereotypic information presented at retrieval is likely to lead to a consistency effect. However it can only be concluded that the present case offers little reason to suppose that stereotypic information acts as a retrieval schema.

Summary and Conclusions

The present experiment was performed with the intention of investigating the biassing effects of stereotypic information presented during the retrieval of material with greater external validity than had previously been investigated. Also, the effects of delay periods comparable to those encountered by witnesses to actual crimes were investigated. The experiment was designed so as roughly to parallel the effects that might be expected were a piece of stereotypic information to be dropped during a police interview about a witnessed crime. Counter to predictions derived from previous research in the area (Snyder and Uranowitz, 1978 a, b, and c; Wyer, Bodenhausen and Srull, 1984; Stangor and Ruble, 1989) stereotypic information was not found to cause a consistency effect in recall. In some cases, stereotypic information presented at retrieval may prove to have the kind of positive effect on memory that it is characterised as having by, e.g., Snyder and Uranowitz (1978c). In other cases, when a slightly different methodology is employed, it may be found to have little or no effect on memory (e.g., Bellezza and Bower, 1981; Rothbart, Evans and Fulero, 1979). Quite possibly, the effects of stereotypic information presented at retrieval will eventually prove as complex and task-dependent as those of stereotypic information presented at encoding (Srull, 1984).

The effects of delay on memory also ran counter to hypotheses. Although a trend was discovered towards deterioration of memory with time, no significant result was in fact discovered. This parallels the finding of Yuille and his colleagues (see above) and contradicts the assumption of many researchers that eyewitness memory is likely to follow an Ebbinghaus decay curve. There was no interaction of delay time with stereotype presentation. It should be mentioned however that the present experiment is subject to the same kinds of criticism as that of Yuille and Cutshall (1986). Both experiments used volunteers drawn from the total set of witnesses to an event. Hence those witnesses with less confidence in their memory for the incident may, as mentioned above, have "deselected" themselves from the experiment. Hence, both Yuille and Cutshall's experiment and the present experiment may provide a more optimistic assessment of eyewitnesses' memory performance over time than is really justified. Therefore, to some extent, both studies may underestimate the effect of delay on eyewitnesses' memory performance.

The present experiment, then, taken with those of Experiments 4 and 5, implies that, although stereotypic information may have some effect on eyewitness memory when presented after a crime, these effects are extremely difficult to characterise in advance, and are likely to be so subtle that, in the complex, real-world situation, they may well be negligible. Although a certain amount of support has

been gained for Snyder and Uranowitz' claim that stereotypes act as retrieval schemata, this effect appears very elusive, and the use of stereotypic information during the police interview cannot be recommended for improving recall. Any distorting effect will likewise probably be quite subtle and will be likely to lead in the direction of a consistency rather than an inconsistency effect. However, the fact that the opposite result was discovered in Experiment 5, implies that the effects may be context-dependent. This is an area which would profit from a great deal more research, particularly with an eye to developing theories of social memory which predict what effect is likely to occur under which conditions. As far as eyewitness memory is concerned, however, the effect of stereotypic information presented at encoding appears to be of more direct relevance, in which case distorting effects appear most likely to occur. The following section of the thesis considers a means by which the impact of such information may be reduced.

CHAPTER 10 - INTERVIEWING EYEWITNESSES: A REVIEW OF THE LITERATURE

> "A critical component of effective law enforcement is the ability of police investigators to obtain accurate and detailed information from eyewitnesses. One experienced judge has stated that incorrect eyewitness identifications have led to more miscarriages of justice than all other factors combined (Sobell, 1972). It is ironic that although the quality of a witness's [*sic*] or victim's report is of paramount importance in solving criminal cases, police investigators often have minimal guidance in developing effective interview techniques to facilitate memory retrieval." (Geiselman, Fisher, MacKinnon and Holland, 1986, pp 385 - 386.)

The previous sections of this thesis have considered ways in which eyewitnesses' judgment or memory for crimes may be affected by stereotypic information. Although some theories (Snyder and Uranowitz, 1978c; Snyder, 1981,) would predict improved memory for stereotype-consistent material with presentation of stereotypic information at retrieval, Experiments 4 and 5 appear to cast some doubt over the validity of this prediction. Therefore it appears likely that the usual effect of stereotypes in the eyewitness situation will be to distort the eyewitness' memory or judgment.

It is therefore important for this thesis to address the question of how such negative effects may be reduced or eliminated. The present section of the thesis reviews the literature on police interview techniques and reports a pair of experiments which aims to develop an interview

technique which can be used to minimise the distorting effects of stereotypic information on memory. These experiments also play an important role in the development of a model of stereotypic effects in the eyewitness situation, to be presented in Chapter 13.

The present chapter consists of a review of the interviewing literature; considering the so-called "standard police interview" (i.e., that performed by otherwise untrained officers), as contrasted with the "optimal police interview", theform of which can be derived from the literature, and techniques which have been developed over the years by psychologists concerned with the area. Particular attention will be focussed on the forensic use of hypnosis, contextual reinstatement, conversation management, and the cognitive interview (basic and enhanced). The rather controversial area of police training in interviewing, (E Shepherd, 1985/91; Roy, 1991; Yuille, 1985,) is not of direct relevance to the present thesis and is therefore not discussed.

The "Optimal Police Interview"

The optimal structure of the police interview can be derived from a study of the literature. Hilgard and Loftus (1979) provide a (somewhat dated) account of the literature which provides the theoretical basis for the structure of such an "optimal police interview".

There is strong evidence that narrative accounts are associated with greater accuracy than responses to direct questions (Cady, 1924, cited in Hilgard and Loftus, 1979; Dent and Stephenson, 1979; Lipton, 1977). This increase in accuracy may be due to the witness producing a smaller quantity of information in the narrative format. This in turn may be associated with the witness adopting a more conservative response criterion in this format (Deffenbacher and Horney, 1981). Comparisons of question-and-answer with narrative formats find an advantage of up to 200% in completeness for the former (Hilgard and Loftus, 1979). Hence, for the sake of production of a greater quantity of information, it is important to follow the free-recall phase of an interview with a question-and-answer phase (Geiselman, Fisher and Raymond, 1987; Loftus, 1979; Roy, 1991).

The specific wording of the questions used in this phase is also important, since, as Loftus has repeatedly demonstrated, slight differences in question form can produce large differences in recall protocols. The effect of misleading questions on the performance of witnesses and their memory trace for the crime remains one of the most heavily-researched areas in the interviewing literature (see, e.g., Geiselman, Fisher, Cohen, Hollander and Surtes, 1986; Loftus, Korf and Schooler, 1988; McCloskey and Zaragoza,

1985; Wright, 1992). Despite this wealth of laboratory work, in actual practice, leading questions remain one of the most difficult interviewing faults to correct. As Cahill and Mingay (1986, p 212) remark: "It is highly likely that a number of leading questions will be asked of witnesses, no matter how assidiously the officer tries to avoid doing so". Perhaps all that can realistically be hoped for in practice is that the negative effects of leading questions be kept to a minimum.

In summary, an optimal police interview would open with a narrative-recall phase, during which the interviewing officer would expect the witness to produce highly accurate but incomplete information. At the completion of the narrative stage, the officer would ask direct questions of the witness, here expecting more complete but less accurate information to be generated. During this stage, the officer should carefully select his questions so as to minimise the chance of producing distortions in the witness' account of the incident.

The following section considers the police interview as it appears to take place in practice.

The Police Interview

Research into the interview process has been an active area for many years in areas such as employment and personnel management (see, e.g.,

Schuh, 1981), but has begun to receive close attention from forensic and criminological psychologists only in relatively recent years (Fisher, Geiselman and Raymond, 1987; Smith and Ellsworth, 1987). Previous research considered relatively small details of the police interview (such as the misinformation effect: e.g., Loftus and Zanni, 1975; Loftus, Miller and Burns, 1978); however, until recently, the basic structure of the police interview seems to have been considered (not unreasonably, perhaps,) the exclusive property of the police themselves.

Child Witnesses: Special approaches (interview techniques, recommendations to the legal system, etc.) have been developed for use with special categories of witness. These include the child witness (Davies, Flin and Baxter, 1986; Dent and Stephenson, 1979; Flin, Davies and Stevenson, 1981; Flin and Tarrant, 1989); child victims of sexual abuse (see Friedemann and Morgan, 1985, for a discussion of the use of the "anatomical doll"); and the mentally handicapped (Tully and Cahill, 1984; Cahill and Mingay, 1986). Indeed, over the last several years, more work has addressed the issue of the child witness than that of any other special category of witness. A great deal of this research has addressed the question of whether or not children are particularly poor eyewitnesses and whether their testimony should be discounted on this ground. However, as Flin (1988) remarks,

since most eyewitnesses are not perfect anyway, the difference between a child and an adult eyewitness is more likely to be quantitative than qualitative. The most interesting work in the field has accordingly addressed the degree of difference between adult and child witnesses in, for example, interrogative suggestibility and courtroom performance. Flin, Davies and Stevenson (1987) present a study of the issues of competence, credibility, and stress in the child eyewitness situation which exposes the then-current limitations of research in the area: for example, little research had addressed the honesty of children, or their ability to use the important investigative tool, Photofit. More recent research has helped fill the gaps in the literature. Brooks and Siegal (1991), found that children varied in their ability to remember real-life events, but were best at remembering events with some personal significance. The suggestibility of children - long a bone of contention in the field - was found not to be fixed but to vary with the environment. They conclude that children of all ages, if properly prepared and handled, can give accurate and informed evidence, (see Baxter, 1990, and Yuille, 1988, for similar arguments; and Kaplan, 1990, for a cautionary note regarding false reports which may occasionally be given by children). A more important question may be the extent to which

jurors trust children's testimony, regardless of how reliable it can be shown to be. Davies, Flin and Baxter (1986), and Flin, Davies and Stevenson (1987), present the intuitively appealing conclusion that most people will tend to believe the testimony of an adult over that of a child. This effect appears to be mediated by perceptions of the confidence and consistency of the witness. Moston (1989) shows that the effect may be due to the way in which children present their evidence: using a great deal of "powerless" speech; a habit which reduces the credibility of adult witnesses. This tendency may be exacerbated in the intimidating environment of the courtroom. It is an important goal of future research to establish which aspects of the courtroom experience children find most stressful. It is possible that children's lack of knowledge of the criminal justice system (Flin, Stevenson and Davies, 1989; Flin and Tarrant, 1989) will prove a particularly important factor here.

Adult Witnesses: The present section considers the "standard" police interview involving adult eyewitnesses of normal intelligence.

Fisher *et al* (1987) provide descriptions and critiques of real interviews conducted by experienced detective officers of the Robbery Division of the Metro-Dade, Florida, Police Departments over a period of four months. A

typical interview was found to begin with a single open-ended question asking the interviewee to recall as much as possible about the event. From that point on, the interview consisted of a narrative account by the eyewitness, punctuated by a series of direct short-answer questions about specific details of the event. Fisher and Price-Roush, (in an unpublished study cited in Roy, 1991), find a number of problems associated with this interview format. One of these is the inappropriate sequencing of questions. Questions dealing with topics that are "perceptually related" to each other are often split by questions dealing with "perceptually unrelated" topics: e.g., two questions about a suspect's eyes may be split by an intervening question concerning his shoes. The presence of such an intervening question may reduce accuracy by approximately 15% (see also Bekerian and Bowers, 1983).

Even more important, perhaps, is the excessive use by interviewing officers of the question-answer format. Thirty-five per cent of all correct statements may be generated within the brief, initial free-recall period of an eyewitness' account (Fisher *et al*, 1987). However, the narrative portion of an eyewitness' statement is interrupted, on average, every 7.5 seconds by the police interviewer. As E Shepherd (1991) remarks, police officers "reduce the investigation to a *questionnaire-like* scenario" (p

4, his italics) - a most ineffective interviewing technique, particularly when the actual case does not fit the preconceived scenario (Mortimer, 1991: this issue is discussed more fully in Chapter 6). Fisher *et al* found that no witness in their study was allowed to complete the free-recall part of the description without interruption. This causes disruption of concentration, and the tailoring of responses to fit the expected interval between interruptions.

In the average interview, Fisher *et al* found the ratio of closed- to open questions to be almost as low as 9:1 (see also Schmitz and Plate, 1978). The sequencing of questions was often found to be incompatible with the eyewitness' mental representation of the facts (e.g., an eyewitness is thwarted in the attempt to recount events in chronological order by a police officer who insists on asking questions that jump around in time).

Roy (1991) provides a brief list of further bugbears in the standard police interview. These are: negative phraseology; non-neutral wording; inappropriate language (e.g., to the eyewitness' mental abilities or social group); staccato style of questioning; judgmental comments; lack of follow-up of potential leads; and under-emphasis of auditory cues. All of these faults could easily be rectified by the briefest of training;

yet they continue to exert negative influence over the efficacy of police interviewing.

The "human element" of the police interview is emphasized by Koepp, Lubbers, Lucero, Mankiewicz and Mason (1981). Neglect of the human element is a major factor in reducing the efficacy of the police interview. The interview should take into account such things as how the emotions of the witness are likely to be running (e.g., s/he is likely to be feeling nervous, apprehensive, and generally ill-at-ease) and the peculiar point of view (both physical and psychological) from which the witness perceived the crime in question. Schmitz and Plate (1978) point out that the interviewing officer is naturally placed in a position of superiority over the witness. A situation which is socially uncomfortable for the witness may further affect his/her performance. S/he may become perplexed, and fail to understand exactly what is required. This can be exacerbated by an officer who sees interactions with the public in terms of dominance-submission (E Shepherd, 1985).

Smith and Ellsworth (1987) found that the way in which an investigating police officer presents himself to the witness affects that witness' response to any information that the officer might provide. When the interviewer presents himself as more rather than less knowledgeable about the case in hand, questions which contain misleading

information produce higher error rates in responding. This factor appears unlikely to be controlled for in the usual police interview situation. Indeed, the importance of appearing knowledgeable is emphasised in some texts dealing with police interviews:

> "...if the encounter between superior and subordinate is about mental rather than physical resources, then the superior must be able to outwit his subordinate. To display mere physical superiority in such a situation is the first step towards loss of status." (Irving and Hilgendorf, 1980, paraphrasing Morris, 1971)

Both Firth (1975) and Goodsall (1974), in their discussions of the police interview, emphasise the need for the interviewing officer to keep talking and appear knowledgeable about the case in hand. As E Shepherd (1991) remarks, however, it may be virtually impossible to disguise lack of knowledge about a case from a witness or a suspect: and the attempt is most likely simply to produce "resistance" in the interviewee.

The attitudes of the witness can also affect the results of the interview. Boon and Davies (1987a) suggest that the accuracy of an eyewitness' memory could be enhanced where only free recall (specifically without prompting by the investigating officer) was employed. This is reminiscent of recommendations made by Hilgard and Loftus (1979), and Fisher *et al* (1987). The attitude of the eyewitness to the police officer himself will also tend to affect how much

information s/he gives. Some eyewitnesses are very uncooperative (E Shepherd, 1991).

Some faults lie neither with the interviewing officer nor with the witness exclusively. Palumbo (1975) points out that a large number of police interviews are conducted in surroundings that are not beneficial to either interviewer or interviewee. Often, they are conducted in distracting or stressful environments, when they could quite easily be held in quieter and more controlled surroundings: e.g., the witness' own home or place of work (although see Leonard, 1971). The less ill-at-ease the witness feels, the better his/her memory performance is likely to be. It is unnecessarily detrimental to the success of the interview for it to be conducted in surroundings that are likely in and of themselves to make a witness feel ill-at-ease.

It seems likely that the "standard police interview" would benefit from eliminating the faults highlighted in this section, and more closely approaching the "optimal interview" outlined above. This could be achieved through more effective training in interview techniques, or application of specific techniques to improve the quality of information elicited. The use of special interview techniques is discussed below.

Techniques to Aid Retrieval

Even to the layman, the use of retrieval techniques must appear a promising avenue for exploration in the interview situation: in everyday life, ordinary people can and do utilise strategies to aid retrieval of information from memory, whether they are aware of them or not. Williams and Hollan (1981) offer a model of long-term memory retrieval as a problem-solving process, involving at least four separate mnemonic strategies. The application of this process appears to result in considerable, if sub-optimal, gain in accuracy. Another everyday example is that of context-dependent memory: it is far easier, for example, to recall the members of a particular jazz band when listening to jazz music than when listening to One FM. Such effects have been experimentally demonstrated in a number of studies, e.g., Godden and Baddeley (1975), Smith (1979). Many other techniques have been investigated in the literature, using a range of methodologies and target materials. Loftus and Marburger (1983) used "landmark events" to help subjects temporally to localise events in their personal histories ("Since the eruption of Mt St Helens, has anyone beaten you up?"). Anderson and Pichert (1978) demonstrated the utility of adopting different perspectives on memory for a prose passage which described a house. Different items of information about the house were recalled

with a change in retrieval perspective - subjects were asked to adopt the perspective of a burglar or a prospective home buyer. Bekerian and Bowers (1983) considered the memory advantage to be gained from the non-random presentation of questions about an eyewitness situation. They exposed subjects to a target event and tested memory for the event, presenting questions in either a random order or in the same order in which the even occurred. When questions were presented in the same order as the original event, subjects were more accurate then when questions were randomly presented. Sobell, Toneatto, Sobell, Schuller and Maxwell (1990) considered the use of memory aids on the reduction of errors in life event inventories, and found that the use of such aids reduced the number of dates reported incorrectly, gave more reliable reports (as measured across two interviews), and led to the recall of *more* events. Sheikh, Hill and Yesavage (1986) found that pretraining, plus the use of imagery-based mnemonics, was beneficial for patients suffering age-associated memory impairment. (See also Patten, 1990, for a history of the use of mnemonics in encoding and at retrieval). This section considers the application of specific retrieval techniques to the eyewitness situation.

Hypnosis

Hypnosis has been conceptualized as anything from "believed-in imagining" to highly focussed attention (see e.g., Orne *et al*, 1984). It is perhaps best defined by its characteristics, which are unlike those of any other state. It is a state in which subjects characteristically forgo evaluation of their experiences; leaving this in the hands of the hypnotist. It is the fact that a hypnotised subject can be requested to relive past experiences that makes its potential for memory-enhancement so seductive.

The use of hypnosis in the eyewitness situation was first proposed in an article by Arons in 1967. Since then, much anecdotal evidence has been provided for its efficacy (see e.g., Reiser 1974, 1976). However, "These field studies are more a reflection of the benefits perceived by those administering the procedure than a more objective evaluation requiring documented corroboration" (Timm, 1982, cited in Orne *et al*, 1984). It must be said that the efficacy of hypnosis has been much questioned in recent years, (see, e.g., Cooper and London, 1973; Dywan and Bowers, 1983; Putnam, 1979; Yuille and McEwan, 1985). Laboratory support for the efficacy of hypnosis has been extremely slight. Many experts now hold hypnosis to be only rarely a useful technique for memory-enhancement; it has even been suggested that, in the cases where hypnosis *has* allowed the

recall of previously unrecallable material, this was caused not by hypnosis *per se*, but by the specific types of question asked of the hypnotised witness; types which contain known memory aids (Yuille and Kim, 1987).

Orne *et al* (1984) provide a very good review of the hypnosis literature, and the following section is partially endebted to their work. As mentioned above, few laboratory investigations have found hypnosis to be an effective memory aid. A common means of using hypnosis as a mnemonic involves age-regression: whether to childhood or just a few days or months previously. The subject is asked to imagine that s/he is once again at the age at which the events in question occurred. S/he is asked to answer questions about the event. Very few studies have found this approach to be effective; and those that have appear to be methodologically flawed, e.g. Reiff and Sheerer (1959), True (1949), in which hypermnesic effects were found to be caused by experimenter bias. It is possible that hypnosis might be useful where stress was encountered at encoding, but the evidence for this is slim. Only one study (Rosenthal, 1944) has found this effect.

The most important type of study for the present case however is that in which the memory of hypnotised subjects for verifiable information is compared with that of non-hypnotised subjects. This experimental paradigm allows for rigorous

control of materials and procedure, and for meaningful comparisons to be made across paradigms.

The literature here is surprisingly confused. However if one follows the example of Orne *et al* and considers studies which have used meaningful material separately from those which have used non-meaningful material, the picture becomes clearer. Hypnosis appears to provide no memory enhancement for non-meaningful material, but it does aid memory for meaningful material - at least for "some individuals in some situations" (Orne *et al*, 1984, p 185).

Most relevant here of course are those experiments in which memory for mock crimes has been tested. Only one such study (Griffin, 1980) appears to have found evidence for hypnotic hypermnesia. All other studies, including better-designed studies such as that by Putnam (1979), have found none.

The most positive results appear where the format for remembering is somehow structured, rather than when the subject is simply asked freely to recall everything that s/he witnessed. Unfortunately, such a paradigm has few implications for the real world. Before one can structure the memory format for recall of an incident, one must have some knowledge of that incident. Yet it is precisely when least is known about the event in question that hypnosis is most

useful to the investigator. Additionally, the small increases that have been seen in the amount of correct information produced by hypnotised subjects are not necessarily meaningful. It is possible that hypnosis simply acts to lower the subject's response criterion: i.e., it encourages him/her to report more information, independent of correctness. Of course, when an investigator is simply trying to generate leads in a criminal case this need not be a great problem: however, later in the investigation, accurate information may be at a premium. Unfortunately, experimenters in the area have tended to measure only the accuracy of memory (i.e. number of correct items of information divided by total number of items generated), rather than response criteria.

The only study to have taken note of response criterion shifts appears to be that by Dywan and Bowers (1983), who found that apparent hypnotic hypermnesia was in fact due to adjustment of subjects' response-criterion. In a similar experiment, Stager and Lundy (1984) used a probed-recall paradigm, and appeared to elicit genuine hypnotic hypermnesia. However, again, probed recall is hardly a useful strategy for the police officer who is trying to find out as much new information as he can about a case in hand.

In conclusion, laboratory evidence for the efficacy of hypnosis as a retrieval mnemonic is at best mixed and at worst extremely poor. There is

little empirical justification for the use of hypnosis in the crime investigation situation: at least at stages of the investigation at which correct information is necessary. As a last-ditch attempt to generate new leads in a stagnating case, the use of hypnosis may have more justification; even then however it should be borne in mind that the information generated should not be relied upon with great confidence.

In addition to the evidence of the laboratory studies, there are further theoretical and practical drawbacks to the use of hypnosis in the eyewitness memory situation.

On the practical side, the safe use of hypnosis requires the presence of a trained hypnotist: and training is costly and time-consuming. Also, hypnotic induction itself is time-consuming: Geiselman, Fisher, MacKinnon and Holland (1985) found that induction took a mean of 27.1 minutes. Since police time is limited and expensive, a less time-consuming strategy would obviously be preferred. Also, hypnosis cannot be used on all subjects. Weitzenhoffer (1958) estimates that 78 - 97% of people can reach the first level of hypnosis, but a deeper level than this is needed for regression. Indeed, up to 5% of people may be totally unhypnotisable.

There is also the problem of increased suggestibility under hypnosis. Putnam (1979) reports the only study to deal exclusively with

this phenomenon. After viewing a videotape of a motor accident, subjects were questioned in either a waking or a hypnotised state. Hypnotised subjects made significantly more errors; and this was independant of the interval between encoding and retrieval. They were found to be more confident in their answers, and believed themselves to be more accurate. This suggests that hypnotised subjects may have had perfectly clear visualisations of material which they confabulate. Increased suggestibility may be due either to shifts in the response criterion or to attempts to please the hypnotist, (an effect discussed at some length by Orne *et al*, 1984).

Finally, the theory generally supposed to underpin the use of hypnosis is flawed. Most advocates of the use of hypnosis (e.g., Reiser, 1974) claim human memory to resemble a video-recorder in that it records any information that enters it, and that what is recorded always has the potential to be played back, albeit only when the subject is under hypnosis. This theory is not currently accepted. Today, memory is held to be more of a reconstructive process than investigators like Reiser would have it.

These problems with hypnosis led to its use being outlawed in many American states. Other strategies may appear more promising.

Contextual Reinstatement

> "... the effectiveness of a retrieval cue is related to the amount of feature overlap with the encoded event (Flexser and Tulving, 1978). This is a statement of encoding specificity (Tulving and Thomson, 1973). One way to maximise feature overlap is to reinstate the context that surrounded the incident." (Geiselman, 1988, pp 245 - 246)

Thus stated, contextual reinstatement appears a simple and elegant means of improving eyewitness memory. One of the earliest attempts to utilise the technique to this end was in an experiment by Malpass and Devine (1981b), who used a contextual reinstatement paradigm which they termed "guided memory". Subjects were interviewed five months after witnessing a staged vandalism in which laboratory equipment was smashed. They were then asked to identify the vandal from an identification parade. Half of the subjects received a guided memory inteview, in which both internal contexts (emotional, etc.) and external contexts (physical features of the room in which witnessing took place) were reinstated. The other subjects acted as controls. Those who received the context-reinstating interview were found to be significantly more accurate in identifications than were controls.

However, the application of contextual reinstatement to the present area is somewhat problematic. Indeed, the very concept of context is difficult to define. As Thomson and Davies

(1988) point out, for some researchers, it refers to "the setting in which a target is found"; whereas for others it is assigned a more active role: "it is the context which determines the figure or target" (Thomson and Davies, 1988, p 2). The literature on the issue is remarkably heterogenous: the contexts which have been manipulated range from discrete words to physical settings to the emotional state of the subject; targets range from words to faces to pictures of objects. The experimental tasks and dependent measures used also vary greatly. Indeed, Thomson and Davies remark that "[i]t is tempting to dismiss context as a useful concept" (p 4).

McSpadden, Schooler and Loftus (1988) remark that recent research has proven ambiguous with regard to the usefulness of contextual reinstatement, and report three experiments which "painted a rather messy picture of context reinstatement" p 226). Geiselman (1988) remarks that contextual reinstatement on its own may not be a useful tool for memory enhancement. In light of the very diverse findings reported in the area, this is a tempting conclusion to endorse. Fernandez and Glenberg (1985, cited in McSpadden *et al*, 1988) report a series of experiments which fail to demonstrate reliable benefits with the use of contextual reinstatement. Geiselman's own work with his colleagues on the Cognitive Interview (see the separate sub-section, below), of which

contextual reinstatement forms a principle mnemonic, however, does indicate that the technique is useful, at least when used in tandem with other techniques (see especially Geiselman, Fisher, MacKinnon and Holland, 1985).

More recent studies have utilised a Retroactive Interference design, by which one set of material is learned in environment *x*, and a second set in environment *y*. Memory for the first set of material is then tested, in either environment *x* or environment *y*. Using this design, positive effects have frequently been demonstrated (e.g., Bilodeau and Schlosberg, 1951; Greenspoon and Ranyard, 1957; Zentall, 1970, all cited in Godden and Baddeley, 1975). However, this paradigm is not without its critics. Strand (1970) suggests that the effects discovered in these experiments were not caused by testing environment *per se*, but by the disruption caused when a subject moves from one environment to another. With this variable controlled, she claims, there may be no effect of contextual reinstatement. However, Smith *et al* (1978) take issue with this claim, presenting experimental data which cannot be accounted for by this disruption hypothesis.

Certainly, the experimental evidence for the efficacy of contextual reinstatement appears to be less than fully convincing. However, as Geiselman points out, the published experimental evidence tends on the whole to be more positive than

negative (although see McSpadden *et al* for a rejoinder to this claim). Perhaps the most thorough set of investigations into the effect is that which has been presented by Cutler, Penrod, and their associates (Cutler and Penrod, 1988; Cutler, Penrod and Martens, 1987 a and b; Cutler, Penrod, O'Rourke and Martens, 1986; Krafka and Penrod, 1985; Shapiro and Penrod, 1986): these studies are reviewed in some detail in Chapter 4. This set of experiments do appear to show that the effect of contextual reinstatement may in many instances be more positive than the brief survey of the literature above might lead one to believe, and that, especially when used in tandem with other retrieval techniques (as recommended by Geiselman, 1988), may prove a useful tool in police investigations.

Conversation Management

The idea underpinning conversation management (CM) is that even in the absence of specific cognitive techniques for improving the quality of memory retrieval, improved quality of interviewer-interviewee discourse can improve the quality of information generated, and reduce interviewee resistance. Of course, this idea is not markedly unique. As long ago as 1975, Firth recommended interviewing officers to use many techniques which are part of conversation-management.

Under CM, police officers are trained to use conversational techniques to create new situations, or alter existing ones, in such a way as to build a "psychological bridge" (E Shepherd, 1991) with the interviewee. The first step in achieving this is to train police to break away from their "traditional" perspective of seeing encounters with the public in purely professional - as opposed to interpersonal - terms. Those undergoing training are encouraged to combine a controlling influence on the interview with a conversational attitude which is socially rewarding for the interviewee.

In each interview, the police officer should invoke the GEMAC script. This outlines a series of stages through which a successful interview should pass:

1. Greeting - which sets the tone of the interview, and provides the interviewee with information as to the officer's attitude *vis-à-vis* the interview.

2. Explanation - which provides a "route map" for the interview. E Shepherd (1991) argues that when no route map is specified, police officers not only indulge in non-systematic, haphazard questioning, but also limit their ability to monitor changes in an interviewee's non-verbal communication as each new topic is broached.

3. Mutual Activity. When the officer reviews the accounts presented by the interviewee, the

interviewee is able to check, alter, or supplement his/her account of the event. At this point the conversational skills of the interviewer come into their own: it is necessary for the him/her to try to develop the "interview spiral", or systematic opening and exhaustion of topics, and probing of both verbal and non-verbal responses.

4. Close - the officer signals the end of the interview.

GEMAC gives a practical framework for structuring an interview using both cognitive and social psychological processes that require what Roy (1991) refers to as "personal micro-skills" on the part of the interviewer, e.g., observations of and memory for the actions of the interviewee; listening and assertion; appropriate questioning, etc. Many of the factors involved in these "micro-skills" appear obvious: e.g., under "listening and assertion" it is specified by E Shepherd (1991, pp 5 - 6) that the interviewer should not speak while the witness is trying to speak. The use of these micro-skills allows the development of "macro-skills", which may be of more direct use to the police interviewer. These include the ability changes in verbal and non-verbal responses, and to detect emotional, motivational, attitudinal, and dispositional changes in the witness.

Use of CM is intended to lead to a more flowing conversation, and the production of a more

accurate and well-structured account of a crime. Indeed, many of the points of CM are applicable and likely to be of advantage whatever other interview technique is employed. George (1992) has tested the efficacy of CM used in conjunction with the cognitive interview, as compared with CM alone, the cognitive interview alone, and the "standard interview". He found a significant interaction between CM and the cognitive interview on the amount of information obtained per question, whereas the effects of either technique alone were non-significant. The same pattern of results was found for the amount of information generated relevant to the perpetrator of the crime.

Although this study implies that CM is maximally effective only when used in conjunction with another retrieval technique, it does appear to have won the approval of the British police. Despite the costliness of training recruits, (George, 1992,) by 1991, twenty-six police forces throughout the UK had adopted the technique (Roy, 1991, p 29). Even such minimal training as a single day spent learning "communication skills" appears to be effective to some degree (J Adams, 1985).

The Cognitive Interview

The cognitive interview has had two incarnations: basic (CI) and enhanced (ECI). Each consists

essentially of four general memory techniques which accord with two principles of memory function: 1. that "the memory trace is composed of several features, and that the effectiveness of a retrieval cue is related to the amount of feature overlap with the encoded event", and 2. that "there may be several retrieval paths to the encoded event" (Geiselman, Fisher *et al*, 1985, p 402). The techniques are discussed briefly below. The first two accord with the first "principle of memory function" outlined above; the second two with the second.

Using the first technique, the interviewer asks the witness mentally to reinstate the context (both internal and external) which existed when the crime was witnessed. A similar "jogging" of memory should be achieved under these conditions as when a witness returns to the scene of the crime; without the problems associated with actually doing this (George and Clifford, 1991). Memon and Köhnken (1992) believe this to be the most effective component of the CI.

The second technique involves instructing the witness to recall the witnessed event from a variety of physical perspectives: s/he attempts to adopt the physical position of, say, the victim, or another witness, and describes what that person would have been able to see. This technique is reportedly most useful when dealing with child witnesses, where the child is asked to put

him/herself into the body of another person, and ask what that person saw (Saywitz, Geiselman and Bornstein, 1992).

The third technique is for the witness to report everything that s/he can remember, including partial information. This may be effective for two reasons: 1. a witness may report only what s/he considers to be relevant, but may not know precisely what information is of investigative value, and, 2. recall of partial details may lead to subsequent recall of additional relevant information (Geiselman, 1988).

Finally, once the witness has provided an account of the crime, s/he is asked to recall the events a second time, but in a different order. The witness may begin at any point: at the beginning, middle, end, or simply the most memorable part of the crime; although the usual procedure is to ask the witness to recall the event first in chronological and then in reverse order (Memon and Köhnken, 1992).

The CI also contains supplementary techniques for the generation of specific items which were not recalled during the narrative phase of the interview, e.g., for the physical appearance of the suspect, asking "Did he remind you of anyone you know?" (Memon and Köhnken, 1992, p 40).

Other studies have extended research into the cognitive interview to the forensically important area of memory for licence plates; including

special menmonics to deal with the particular problems associated with this (MacKinnon, O'Reilly and Geiselman, 1988; Mende, MacKinnon and Geiselman, 1986).

The first empirical study of the efficacy of the CI was reported by Geiselman *et al* (1984). Undergraduate subjects were interviewed about a staged crime, using either the CI or standard police interview. Under the CI, significantly more correct information was generated for two of three classes of memory (for persons and events, but not for objects), with no significant increase in the amount of incorrect information generated. Geiselman *et al* (1985) compared both the CI and hypnosis with the standard police interview on memory for a violent police training film used by the Los Angeles Police Department. Both the CI and hypnosis were found to elicit approximately 35% more correct information than the standard interview. The number of incorrect and confabulated details did not differ significantly with condition. A subsequent study (Geiselman *et al*, 1986) showed that the hypermnesic effect of the CI did not appear to be limited to any particular class of witness, but to be generalizable across age, ethnic group, annual income, and education level. These findings have been replicated in Germany (Aschermann, Mantwill and Köhnken, 1991, cited in Memon and Bull, 1991).

Perry and Chapman (1992) used the CI to study memory for road accidents, finding that the CI produced 15% more correct information than the standard police interview, without increasing the amount of incorrect information generated. Memon and Köhnken (1992) speculate that the comparatively small effect discovered here is due to the unusual experimental procedure employed: Half of the subjects were interviewed under standard interview conditions after they had already experienced CI techniques one week earlier, in the first part of the experiment. These subjects may have applied CI techniques even in the standard interview. Newlands (1993) used a video-recording of a mugging event to compare memory under CI and standard police interview conditions for recall of specific types of information: for people, surroundings, conversation, actions, and perpetrator information. Significantly more correct information was generated under the CI; however the type of detail which benefitted was information about the surroundings: perhaps the least useful in a genuine crime investigation. A significantly greater number of incorrect details was also generated in the CI than in the standard interview; a finding at odds with those cited above.

The enhanced cognitive interview (ECI) was produced by Fisher, Geiselman, Raymond, Jurkevich

and Warhaftig (1987b), and Geiselman and Fisher (1988), in response to their review (Fisher *et al*, 1987a) of practical problems encountered by police officers who conduct interviews on an everyday basis. The ECI includes three techniques reminiscent of conversation management. These are: rapport building with the interviewee; transference of control to the witness to give him/her time to concentrate and search his/her memory; and structuring of the interview so that it is compatible with the mental operations of the witness - avoiding a rigid, uniform style of questioning. The ECI is therefore a more effortful and time-consuming procedure for the interviewer. Fisher, Geiselman and Amador (1989) recognise that it may be difficult for interviewers to employ these techniques in the absence of proper training. Nevertheless, the ECI appears to be even more effective than the original procedure (Fisher *et al*, 1987b; Geiselman and Fisher, 1988).

Fisher *et al* (1987b) compared the CI with the ECI in a laboratory study of memory for video-recordings of simulated crimes. The ECI was found to generate 45% more correct information than the CI. Since the CI had already been found to elicit 35% more information than the standard police interview, this represented a dramatic effect. This was not found to be caused by the extra time involved in application of the ECI.

George and Clifford (1991) tested the ECI against the standard police interview and ECI-plus-conversation management for recall of a staged incident. Subjects were questioned by experienced detectives, two weeks after the staged incident took place. The ECI alone generated 36% more correct details than the standard police interview; and the ECI with conversation management 42% more correct information again: with no increase in the number of errors made. This is a similar finding to that reported by George (1992), above.

A number of studies of the ECI have been conducted in Germany by Köhnken and his colleagues. Köhnken, Thürer and Zoberbier (1992) exposed subjects to a video-recording of a blood donation. One week later, memory was compared under the standard (or "structured") interview and under the ECI. The ECI was found to give 52% more correct information than the "structured" interview, with no increase in the amount of incorrect information generated. Using similar stimulus materials, Köhnken, Schimossek, Aschermann and Höfer (1991) found the ECI to increase correct recall by 35% over the "structured" interview, although a marginally significant increase was also discovered in the amount of incorrect information generated.

Mantwill, Aschermann and Köhnken (unpublished manuscript, cited in Memon and Köhnken, 1992),

again used the same stimulus materials. Half of their subjects had prior experience as blood donors, and half did not. Interviews were performed by subjects from various professions (none psychologists), all of whom had donated blood. Half of these used the ECI, and half used the "structured" interview. The ECI was found to generate 25% more correct information than the "structured" interview, without any increase in incorrect information. The performance of experienced and inexperienced subjects was not found to differ significantly.

Only two field tests of the ECI appear to have been reported. Fisher, Geiselman and Amador (1989) studied use of the ECI by trained and "control" police detectives from Miami, Florida, when questioning victims and witnesses of actual crimes. The ECI was found to improve interview performance (in terms of pieces of correct and incorrect information generated) both within subjects (before vs. after training in the techniques) and between subjects (ECI interviewers vs. controls). Detectives elicited 48% more correct information after training in the ECI than before training, and 63% more information than the untrained detectives. George and Clifford (1991) used for subjects thirty-two British police officers either trained or untrained in the ECI. The ECI was found to elicit 14% more correct information than the no-training condition. A

before/after measure was also taken. It was found that training in ECI gave a within-subjects increase of 55% in the amount of correct information generated.

A number of recent studies (Geiselman and Padilla, 1988; Saywitz, Geiselman and Bornstein; Saywitz, Geiselman and Bornstein; Köhnken, Thürer and Zoberbier; Memon, Cronin, Eaves, Bull and Küpper - all 1992) have considered the use of the ECI with child witnesses. Although this is an important area of research, this literature is not reviewed here, since it lies some way outside the present field of enquiry.

Hence empirical findings appear to lend great support to the cognitive interview (whether basic or enhanced): "[t]he memory enhancing effect of the CI proved to be extraordinarily consistent and robust as far as adult witnesses are concerned" (Memon and Köhnken, 1992, p 46.) Positive effects appear to hold for a large range of interviewers and interviewees from the USA, the UK, and Germany, and for a number of stimulus materials (unpublished research by Fisher and Quigley, cited in Memon and Bull, 1991, used the CI in an epidemiological study of outbreaks of food-borne disease). Memon and Bull summarise the results of investigations into the CI as showing that it is 20 - 40% more effective than the standard police interview, and that this is independent of the specific interviewer (whether student or police

detective). However one should not be without reservations. As Memon and Bull (1991, p 302) remark, it is unclear how each aspect of the technique works and whether any one mnemonic is more effective than others: of particular concern may be the effects of the "reinstate context" mnemonic. The exact effects of contextual reinstatement on eyewitness memory have yet to be fully specified. The finding reported by Newlands (1993) that use of the CI increased the amount of correct information generated, but specifically for the *least relevant* aspects of the crime situation, should also be investigated. In general, however, research indicates that use of the CI givens positive results, and it appears a most effective technique for the improvement of eyewitness memory.

Conclusions

Eyewitness memory is the single most important source of information in the criminal investigation. Criminal investigation bodies tap eyewitness memory through the use of the interview. Hence, interview techniques that increase the amount of correct information generated - with no concomitant increase in incorrect information - are extremely valuable. While hypnosis is such a controversial technique in this respect that its use has been banned in several states in the USA, both conversation

management and the cognitive interview appear to be more useful. By far the greater amount of research has addressed the cognitive interview. The results of this research have been generally very positive, although a number of criticisms can be made. Perhaps particularly exciting is work by George and Clifford (1991), and George (1992), in which CI and CM techniques have been combined, their positive effects appearing to be additive.

The aim of the present section of this thesis is to develop an interview technique by means of which the potential negative influence of stereotypic information on eyewitness memory might be reduced. While CM techniques might be helpful in reducing the likelihood of stereotypic information being presented at retrieval, the effects of such information are likely to be most pronounced when it is present at encoding. For example, someone may witness a bank robbery, already possessing a stereotype of the type of person who commits bank robberies. Neither CM nor the CI appears well-equipped to counter the effects that are likely to occur under such conditions. No discrimination is made in these techniques between what *is* true, (i.e., what actually happened in the witnessed bank robbery,) and what the witness *believes* to be true, (i.e., the stereotypic information in terms of which s/he encoded the event in the first place). Indeed, no such discrimination is meant to be made. It is,

however, necessary in the present situation. Aditionally, the aims of CM and the CI make them somewhat inappropriate to the present situation. The aim of both techniques is to increase the amount of correct information generated while not increasing the number of errors made. The interview technique needed in the present situation is in some ways opposed to this. It must aim rather to decrease the amount of incorrect information generated while not affecting the amount of correct information. The next two chapters outline such an interview technique, and go on to test it in the laboratory.

CHAPTER 11 - CUT ON THE BIAS: USE OF THE SOURCE-MONITORING TECHNIQUE TO REDUCE STEREOTYPIC DISTORTION IN RECALL

> "'I don't usually...wait a minute, though. I've just remembered. I did rob a bank yesterday. I completely forgot.'"
> David Nobbs: *The Return of Reginald Perrin*

Introduction

The present thesis has so far considered many aspects of the relationship between stereotypic information and eyewitness judgment and memory. Areas which have received particular attention are the effects of stereotypes on judgments (both of actions and their basis, in Experiment 1, and of features of a to-be-remembered target individual, with respect to both face and whole-body information,) and the effect of stereotypes on memory, particularly with respect to the question of whether or not stereotypic information might be employed as a form of last-ditch retrieval mnemonic in police interviews of eyewitnesses. It was argued in the Discussion section of Experiment 6 that stereotypic information appeared unlikely to have any pronounced positive effect upon eyewitness memory, of the sort that was proposed by Snyder and Uranowitz (1978c), and that, therefore, the most pertinent area of study for the real-world eyewitness situation was the reduction of the biassing effects this might have on memory. To this end, the present experiment continues the study of interviews begun in Experiment 6, now considering whether or not it might prove possible to construct an interview by means of which the effect of stereotypic information may be reduced,

where this is deemed by the investigating authority to be appropriate or necessary.

The type of effect under consideration is that mentioned by Shepherd, Ellis, McMurran and Davies (1978), in the quotation cited in Chapter 5. Where an eyewitness' memory of an event is distorted in some way by stereotypic information, likely it will only be under very rare conditions that the stereotypic information and the to-be-remembered information be presented through one and the same medium, or from one and the same source. Far more likely is that stereotypic expectancies or biasses be brought to the situation by the eyewitness, or that such biasses be introduced *after* the event, either during rehearsal in the eyewitness' own mind or discussion with friends, relatives, etc.:

> "When a person witnesses an event, mulls it over in his or her mind, and hears other people talk about it, he or she stores information about the event from all three sources. If later asked to recall the event as it was witnessed, such a person must discriminate memories of the event itself from memories of his or her ruminations and from memories of what other people said about the event." (Lindsay and Johnson, 1989b p 111)

Hence, the central issue here is how to access during the course of an interview only information about what was actually witnessed, and not stereotypic information which may have been presented through other sources and which may bias memory.

Directly relevant here is the work by Marcia Johnson and her colleagues on "source-monitoring". This series of papers represents an attempt to define people's ability to discriminate between information

provided by different sources. Her later experiments - particularly those conducted in partnership with Stephen Lindsay, (e.g., Lindsay and Johnson 1989a, b and c,) - explicity consider the issue of eyewitness memory, and particularly the "suggestibility effect" studied by Elizabeth Loftus and her colleagues. These studies then are of special relevance to the present area of enquiry; however a consideration of them will be preceeded by a brief look at Johnson's earlier studies and the development of the "reality-monitoring" model. The application of source-monitoring techniques to the interview situation is considered in Experiment 7.

Literature Review

Johnson cites an impressive line of descent for her studies, which originally evolved out of an interest in the discriminability of externally- and internally-derived experiences. This family tree includes, among others, Hume, Locke, and Freud (Lindsay and Johnson 1989b); however the earliest relevant studies in the psychological literature are probably those by Perky (1910). Perky had subjects look at a screen while imagining and describing a coloured object. On the screen, he raised to slightly above threshold an image of the object in question. Most subjects were found to report the perceptual stimulus as if it were their imaginings, and were somewhat surprised to be told that they had *seen* the stimulus. This experiment demonstrated an essential similarity between the

imagination and perception which was later to prove a fundamental point in Marcia Johnson's work.

Segal and Nathan (1964) attempted to replicate Perky's results and achieved only a 25% effect: a change in performance levels which they attributed to increased sophistication in subject populations since Perky's day. However, when a nominally more sophisticated apparatus was used for the projection of images, results were found to be similar to those discovered by Perky.

After these papers, the area received very little attention until Johnson took the baton. There follows a brief resumé of her work in the area.

Johnson, Raye, Wang and Taylor (1979) looked at two populations: one of good and one of poor visual imagers. These subjects were either exposed to pictures of common objects, or heard the word corresponding to those pictures and asked to imagine the picture. They were then asked how many times they had actually *seen* the picture. It was found that, for both groups, imaging increased the apparent number of exposures to the picture, and that this effect was particularly pronounced for good imagers.

Johnson, Raye, Foley and Foley (1981) ran a rather more complex study, this time using verbal materials only. This study considered the importance of cognitive operations to the memorability of an item. Two types of item were used: "presented items" and "generated items". For presented items, subjects were shown a category plus an instance of that category

(eg. Animal - Dog). For generated items, subjects were shown a category plus the initial letter of an instance of that category (eg. Animal - D). Two types of generated item were used, *easy* ones and *difficult* ones. An example of an easy item would be Animal - D, since "dog" is a relatively accessible instance of the category "animal"; an example of a difficult item would be Animal - P, since "pig" is a slightly less accessible instance of the category. Subjects were then shown various instances of categories and asked which had been presented and which they had generated themselves. This task was found to be easier for less available instances. It was claimed that the generation of a relatively automatic instance like "dog" was more like perception than true generation. Hence, the experiment shows that the presence of cognitive operations associated with a memory trace can be used as a guide to the origin of that trace. This has obvious implications for studies of eyewitness memory, and particularly Elizabeth Loftus' work on the misinformation effect (see below).

Foley and Johnson (1982) ran an experiment in which the subject and a confederate of the experimenters were presented with a list of words. Of these, the confederate in both conditions had to *say* a random 50%, while subjects had to imagine the other words being said. In one condition, subjects imagined *themselves* saying the words, and in the other imagined the *confederate* saying them. When asked to identify the source of the words, subjects performed more

poorly than they did when they imagined *themselves* saying the words. This supports the hypothesis of the existence of qualitative differences between perceived and non-perceived, imagined events by showing that an increase in feature overlap between the two reduces their discriminability.

Foley, Johnson and Raye (1983) reported two experiments, using child subjects. In one condition, children were asked either to say words or to listen to a confederate of the experimenter saying words ("Say-Listen" condition); in the other they were asked either to say words or to imagine saying them ("Say-Think" condition). They were then given an identification-of-origin test for the words. As predicted, the children made significantly more errors in the Say-Think than in the Say-Listen condition. A second experiment using child subjects showed that a Say-Listen condition was significantly more difficult than a Listen-Listen condition: this is what the reality-monitoring model would predict, since a distinction between memories from two different classes (internal or external) should be easier than a distinction of memories from within one class (external).

Johnson, Kahan and Raye (1984) report a complex experiment in which reality-monitoring for dreams was investigated and compared with that for similar (i.e., dream-like) material generated consciously. Pairs of partners were used for subjects in this experiment. Every night before sleeping each partner was secretly

assigned to one of three conditions; every morning s/he had to report to his or her partner one of the following: a dream *experienced during the night*; a dream which they had *read* the previous evening; or a dream which they had *invented*, using cue-words given them by the experimenter. At the end of this period, the experimenters put together segments of all of these dreams or dreamlike imaginings, and asked subjects to report which segments were their own, which were their partners', and which were new. The reality-monitoring model made the counter-intuitive prediction that it would be easy to discriminate one's own readings or inventions from one's partner's readings or inventions, but difficult to discriminate one's own dreams from one's partners dreams: this prediction on the basis of the level of cognitive effort that has gone into the creation of the memories. This prediction was supported: interestingly, since the experimental support of a counter-intuitive prediction counts for much in the validation of a hypothesis. This lends support to the idea of the discriminability of memories from different sources.

Johnson and Foley (1984) also studied children's memory. They report two experiments. In the first, children were either shown pictures of objects or were read the names of the objects and asked to imagine the corresponding picture. Afterwards, children were asked to judge how many times the pictures had actually been presented. This was found to vary with

the number of times the picture had been imagined as well as with the number of actual presentations of the picture. In the second experiment, discrimination between saying, thinking, and hearing words was examined. In the "Say-Listen" condition, children were asked to say some words and listen to another person saying other words. They were then asked whether the words were a) said by them, b) said by the other person or c) new. The reality-monitoring model predicted poorer performance in the Say-Think than in the Say-Listen condition, since in the former condition there should be relatively little difference in the amount of information that is available about the cognitive operations involved in the creation of the memory. This was indeed found to be the case. Their Experiment 4 involved *action* rather than *verbal* stimuli. The actions involved were simple communicative gestures, extension of a part of the body, looking or orienting towards an object in the room, or actions "performed from a standing position (e.g., stand up, do a jumping jack, and sit back down)" (Johnson and Foley, 1984, p 43). There were three memory conditions: Do-Watch, Watch-Watch, and Do-Think (paralleling Say-Listen, Listen-Listen, and Say-Think conditions in the earlier experiments, respectively). As expected, subjects' performance was more accurate in the Do-Watch than in the Watch-Watch condition. There were no consistent results in the third condition, since in this condition performance was found to vary with age. Once again these

experiments supported the reality-monitoring hypothesis.

Foley and Johnson (1985) continued to investigate children's tendency to make confusions in the Do-Think condition (see above). They found this to generalize from words to simple actions in a Watch-Do condition, in a Watch-Watch condition, or a Do-Think condition.

Johnson (1985) uncovered new evidence that "real" and "unreal" memories were qualitatively different from each other. Subjects attended to both perceived and imagined items. Subjects in one condition were then administered an identification-of-origin task; subjects in the other condition were administered an old/new discrimination task. Johnson found origin of information to interact with the test administered: origin judgments were made more quickly for imagined than for perceived items, whereas old/new judgments were made more quickly for perceived than imagined items.

Suengas and Johnson (1985) decided to try to extend these earlier findings into the arena of more complex, autobiographical "mini-events" (e.g. meeting a Korean student, having coffee and cookies). The next day, subjects were asked to rate their memories along several dimensions. Qualitative and intuitively correct differences were discovered between real and imagined memories, e.g., real memories were sharper, had more colour, more sounds, etc. For real memories, subjects could also better remember how they felt at the time, thought the memory was more revealing about

themselves, and had fewer doubts about its accuracy. On other dimensions, real and imagined memories were not found to differ (e.g., amount of smell, taste, or temperature information). Suengas and Johnson also manipulated the number of times events were thought about (0, 8, or 16 times). Rehearsal affected some but not all aspects of the memories. Generally, rehearsal had parallel effects on perceived and imagined memories. The effect of rehearsal is summarised by Johnson (1987), who claims that rehearsal is selective and concentrates on the whole on *visual* aspects of an experience. It makes these aspects of imagined events less accessible, but makes them more like real events in the amount of cognitive and affective information they contain. Over all, she claims, differences between the two kinds of memory will *increase* with time.

Johnson and Suengas (1989) investigated the spontaneous use of reality-monitoring in the social world. They asked witnesses of certain events to describe those events to a panel of independent judges. They varied the content of the descriptions given by having witnesses rehearse different aspects of the event prior to describing the events. When asked to judge the truth or falsity of these accounts, judges were found to rely on the presence or otherwise of *perceptual* details (cf. Johnson 1987). Obviously, this is consistent with the differences between the two types of memory outlined in the previous research.

Kahan and Johnson (1990) predicted on the basis of the reality-monitoring model a difference in subjects' performance on discrimination task and an old/new differentiation task when similar items were perceived or imagined. Performance on the discrimination task was expected to be poor, while that on the old/new differentiation task was expected to be good. Results were wholly consistent with this prediction.

Johnson (1987) pointed out that reality-monitoring can plausibly be seen as just one aspect of a pervasive feature of memory functioning: i.e., the ability or inability to ascribe *sources* to memories. This general ability she termed "source-monitoring". The discrimination between an internal and an external source - the focus of most of the work up until that time - was one type of source-monitoring; another was the discrimination between two or more external sources ("external source-monitoring"). Obviously, much of the previous work (e.g., some of the conditions employed by Foley and Johnson 1985) had included conditions which necessitated external source-monitoring; nevertheless, this was the first time that the questions had explicitly been raised. It is "source-monitoring" in this more general sense that is of particular relevance in the present instance.

Lindsay and Johnson (1991), following on from this piece of work, demonstrated the existence of a manipulation that had opposite effects on old/new differentiation and source-monitoring ability.

Subjects were presented with nouns for deep or shallow processing. Deep processing was found to improve performance on the old/new differentiation task, but to impair performance on the source-monitoring task, in which they were asked to discriminate between targets and distractors which had also been deeply processed in the experimental situation.

This ability has obvious ramifications for that much-investigated area of eyewitness memory research, the "suggestibility effect" (see, e.g., Loftus, 1977; Loftus and Greene, 1980; Schooler, Gerhard and Loftus, 1986; McCloskey and Zaragoza, 1985, 1987; Wright, 1992). Much subsequent work on source-monitoring has been concerned with investigation of the suggestibility effect. Loftus' claim is that misinformation received after exposure to stimulus material actually alters the memory-trace for that material. There had been much criticism of this "over-writing" hypothesis (see especially Lindsay and Johnson, 1989; McCloskey and Zaragoza, 1985; Wright 1992). Lindsay (1987, cited in Lindsay and Johnson, 1989a) purportedly showed that the eyewitness suggestibility effect could not be wholly ascribed to demand characteristics of the task in question, (as claimed by, e.g., McCloskey and Zaragoza, 1985,) but that genuine memory-source confusions do in fact occur. Schooler *et al* (1986) discuss the suggestibility effect in terms of the Johnson and Raye reality-monitoring model, saying that failures in reality-monitoring may play an important role in

suggestibility. Lindsay and Johnson (1989b, cited in Lindsay and Johnson, 1989c) found that the likelihood of making source-monitoring errors increases with the degree of similarity between potential sources of information, in terms of their perceptual properties, modality of presentation, semantic content, or cognitive operations. Hence:

> "[t]he procedures used in studies of eyewitness suggestibility create ideal conditions for source-monitoring errors. Both the original information and the postevent information concern the same topic, and both are typically presented close together in time, in the same environment, by the same experimenter, and so forth. These similarities may make it difficult for subjects to later discriminate between memories derived from the postevent information and memories derived from the original depiction of the event." (Lindsay and Johnson, 1989c, p 350.)

Suggestibility effects occur, (at least some of the time,) when there is a failure in source-monitoring and subjects mistakenly identify a piece of information as having been presented in the stimulus materials when in fact it was presented in the post-event information, or "summary" of the stimulus materials. It is possible therefore that the suggestibility effect could be at least greatly reduced by the application of source-monitoring techniques to the paradigm. Where memory distortions are caused by stereotypic information, there is an obvious parallel to the misinformation effect. Stereotypic information presented after to-be-remembered material is analagous in its effects to the "misinformation" paradigm in experiments by, e.g., Elizabeth Loftus, causing analagous memory distortions

(see particularly Experiment 3 and Chapter 7). Even where stereotypic information is present at encoding, however, source-monitoring would still be expected to have positive effects: the stereotypic information and the to-be-remembered material could have their origins in different sources, and should therefore be discriminable. Given the similarities between the suggestibility effect and the effect of stereotypic information on social memory (see above), it may be efficacious to apply source-monitoring techniques to the type of paradigm presently under consideration.

Source-monitoring in the eyewitness situation: An initial experiment along these lines was run at the University of St Andrews School of Psychology in the academic year 1991/2 as part of an undergraduate degree. Sussman (1992) showed subjects a video-recording of a meeting in a bar between two actors playing the parts of old schoolfriends, one of whom was dressed smartly and the other untidily. A voice-over presented with the video-recording stated that the smartly-dressed man was a labourer, and the untidily-dressed man a lawyer: this being information counter to previously-established stereotypes of these professions. Subjects were interviewed as to their memory for different types of information (particularly, stereotype-consistent and -inconsistent information) presented (either by the actors in their conversation, or in the voice-over that accompanied the video presentation). In one

condition, subjects were given source-monitoring instructions; the other was a control condition. All interviews had three phases: an initial, free-recall phase, a "readback" phase during which all subjects were read what they had recalled, and source-monitors were additionally asked to attribute a source to each piece of information, and finally a prompt phase in which direct questions were asked of the subjects. It was hypothesized that source-monitors would recall more facts, make fewer errors, and perform more accurately over all; that they would be able accurately to locate the source of their memories; and that they would not be influenced by an inconsistent stereotype. The results of the experiment were however inconclusive. Sussman, in his discussion of the experiment, suggested that the efficacy of the source-monitoring task might be increased by including source-monitoring instructions in the free-recall phase. There were two reasons for this: 1) the procedure should lead to very accurate reports, given the accuracy and volume of information generated on average in the free-recall phase, and subjects' very precise ability to locate the source of information reported, and, 2) the fact that, in Sussman's experiment, source-monitors showed better recall in the prompt phase than did controls, implies either that source-monitoring instructions, to be useful in improving memory, need time to work, or that, given source-monitoring instructions, subjects continue to use these throughout the interview despite only being

specifically *told* to source-monitor before the readback phase. Each of these points gives reason to believe that, the earlier in an interview source-monitoring instructions are given, the more effective the interview may prove to be.

Following this line of reasoning, it was decided to run a conceptual replication of Sussman (1992) which would study the relative efficacy of the source-monitoring task presented with a free-recall component, and a control condition. It was hypothesized that when subjects were presented with the source-monitoring task at the same time as the free-recall component of the interview, then their recall would be more complete and more accurate than that of the other two groups, and less susceptible to the effect of stereotypic information. It was also hypothesized that subjects in the original, "standard", source-monitoring condition would produce more complete and accurate recall, making fewer errors, than subjects in the control condition, but less well than subjects in the source-monitoring-plus-free-recall condition. Subjects in both source-monitoring groups were, in addition, expected to be able to monitor the sources of information which they receive. Consistent with the previous experiment, subjects were expected to show superior memory in the free-recall phase than in either of the two other phases. Both the *amount* of information recalled and the *accuracy* of this information were expected to be greatest during this phase.

Hence, the hypotheses of the present experiment can be summarized as follows:

1. In the initial phase, subjects in the source-monitoring-plus-free-recall condition should remember more information, make fewer errors, and be more accurate than subjects in either of the two other conditions.

2. In the readback phase, subjects in the source-monitoring condition should remember more information, make fewer errors, and be more accurate than subjects in the control condition.

3. Subjects in the source-monitoring conditions should make fewer errors for factual information than subjects in the control condition, and should make fewer errors regarding the true professions of the characters on the video-recording.

4. Subjects should be able correctly to identify the sources of their memories.

5. Overall, performance of subjects in the source-monitoring-plus-free-recall should be superior to that of subjects in the source-monitoring condition (i.e. more correct information generated, fewer errors made, and greater overall accuracy), and performance of subjects in the source-monitoring condition should be superior to that of subjects in the control condition.

Experiment 7

Methodology

Pilot Study: Twelve undergraduate students (six male, six female) took part in a preparatory study to

establish the stereotypes to be used in the experiment. Each subject was asked how they would recognise a member of each of six job categories, judging by their characteristics and behaviour, such as appearance and conversation. The six job categories provided were: doctor, lawyer, labourer, dustman, accountant, and factory worker. Lawyer and labourer were the job-categories with the most clear-cut stereotypes across the twelve subjects, and hence were the categories used in the video-recording.

Subjects: Forty-six students (thirty female, sixteen male) both undergraduates and postgraduates, aged 18 - 27, participated in this experiment for a small fee. One male performed so poorly relative to other subjects (recalling a total of only three items of information) that his results were discounted.

Design: The experiment was a between-subjects design with three interview conditions and fifteen subjects per condition. Each interview had three phases: an initial, a readback, and a prompt phase - although the exact content of each phase varied subtly with condition (see below). The three interview conditions were: control, source-monitoring, and source-monitoring-plus-free-recall. These are referred to as Conditions C, SM and SMFR respectively. The exact contents of each interview type is discussed below.

Materials: A video-recording was made in a local bar, using as principals two amateur actors. They were seated at a corner table in a bar, with a staff exit door visible behind them. On the table was cutlery, drinks, napkins, and an ashtray. Also visible, slightly to one side of the principals, was a girl (in reality also an amateur actress), seated at the adjoining table. The character on the right was wearing a suit, had short hair, and gave the appearance of being more subdued and responsible, whereas the character on the left was wearing a sweatshirt and jeans, had long hair held back in a pony-tail, and gave the appearance of being more self-confident and jovial. Subjects learned that the first of these characters was called Paul, and the second John. The voice-over then stated that Paul was a labourer and John a lawyer. This information was inconsistent with the stereotypes of these professions established in the pilot study. Throughout, the voice-over presented information that was consistent or inconsistent with that presented by the actors. (The script is reproduced in Appendix 31). The complete video-recording was about five minutes long.

Procedure: Subjects took part in the study one at a time. They were seated at a desk at a comfortable distance from a TV set with video-recorder. They were told that they would be watching a video-recording of a meeting between two men in a pub, and that they were

to imagine that they were overhearing, rather than taking part in, their conversation.

At the end of the video-recording, all subjects were administered a three-minute filler task before being interviewed in the various conditions. All interviews were tape-recorded to facilitate later analysis.

There were three phases to each interview. The initial phase was a period of free-recall in which subjects were instructed simply to remember and report everything they could from the video-recording, regardless of its apparent importance. In this phase, subjects in Condition SMFR were asked to tell the interviewer from which of the possible sources (i.e., video or voice-over) they remembered each piece of information they gave. All reported details were transcribed in this phase and read aloud to the subject in the second (readback) phase. In this phase, subjects in Conditions C and SMFR were simply asked to inform the interviewer when and if any piece of information read back to them reminded them of another piece of information which they now remembered but had not reported. Subjects in Condition SM however were asked to tell the interviewer from which source (video or voice-over) they remembered each bit of information. The final phase was a prompt phase, during which subjects in all three conditions were prompted on key points from the video-recording that they had not mentioned. In the case of factual information prompt questions were of the form: "Can you tell me their ages?". For information pertaining

to the conversation, questions were more pointed, e.g., "Do you remember anything about a football match/scout camps?" Factual information could be presented either by the actors, by the voice-over, or both. Some bits of information were consistent across these modalities, and some were inconsistent.

At the end of the prompt phase, subjects were debriefed, paid, and allowed to leave.

Scoring Procedure: Three types of information were provided in the video-recording. These were:

1. Visual information - e.g., descriptions of the setting, clothing of the actors, items on the table, in the background, etc.

2. Conversational information - e.g., points mentioned in the anecdotes told by the two principals.

3. Factual information - facts about the lifestyles and life-histories of the principals, e.g. their families, their ages, their jobs, etc.

Four sub-categories of factual information were used. These were: information appearing only in the video (video-only information), information appearing only in the voice-over (voice-over-only information), or information appearing in both video and voice-over. Items in this last category could be consistent between the two sources (consistent information) or inconsistent between the two sources (inconsistent informaton). For example, the voice-over said that the character Paul had two sons, whereas in the video

he said that he had two daughters. This was a piece of inconsistent, factual information.

Twenty-one points of visual information were available, thirty-six points of conversational information, and nineteen points of factual information (of which two were video-only; eight were voice-over only; six consistent; three inconsistent). Due to the nature of a video-recording and the importance of visual and conversational information in a vide-recording, it was impossible to balance the number of bits of information available in each category.

Each interview was analysed for the type of information remembered in each phase, the type of errors made, and percentage accuracy. Only *new* bits of information were considered. Thus if for example a subject recalled a certain bit of information *x* during the initial phase, and then again during the readback phase, only recall during the initial phase was counted. It was decided to use this method of analysis, rather than scoring each bit of information each time it was recalled, since this approach has more ecological validity in the present field. Police interviewers are more interested in the new bits of information that a witness can provide during an interview than in how many times the witness is able to recall the same bit of information.

Source-monitoring interviews were also analysed for accuracy in locating the source of information. Only answers in the factual information category were

considered here, since only this type of information came from different sources. Within the factual information category, the consistent video/voice-over information was not used, since errors of source-monitoring could not be made for this category of information. If a subject had only learned a bit of information in the voice-over and said this, s/he would not be incorrect in their source allocation as s/he had no knowledge of the information appearing in the video. Inconsistent information was of course presented as a different item via each modality. Hence it was important to check how good subjects were at attributing the correct source to these items of information.

Results

Correct Information:

The data for correct information recalled in each interview group in each phase are presented in Table 11.1, below, and in histogram form in Figures 11.1 and 11.2.

Analysis of variance was performed on the effects of interview group and phase of interview, on the amount and type of correct information remembered. No significant effect of interview group was discovered: $F(2,42) = 2.1229$ $p = 0.1323$. There was however a significant interaction between interview group and phase of interview for information remembered: $F(4/84) = 3.0098$ $p = 0.0226$.

A Neumann-Keuls test on these data established that more correct information was remembered during the

initial phase by subjects in Condition SM than by those in Condition C or Condition SMFR (Q = 3.6 p < 0.01; Q = 1.2 p < 0.01, respectively).

Cond.		Initial	Readback	Prompt	Total
C	M	14.200	1.467	9.933	25.600
	SD	4.998	1.329	4.675	6.563
SM	M	19.933	1.733	6.067	27.733
	SD	4.086	1.234	2.475	4.337
SMFR	M	14.133	1.590	9.667	25.390
	SD	3.932	1.302	2.463	5.040

TABLE 11.1 - MEAN CORRECT INFORMATION REMEMBERED IN EACH INTERVIEW PHASE FOR EACH INTERVIEW TYPE, WITH STANDARD DEVIATIONS

A Neumann-Keuls test also established that there was a significant difference in correct information recalled between the initial phase of Condition SMFR and that of Condition C.

A highly significant effect of phase was also found: F(2/84) = 152.5074 p = 0.0000). A Neumann-Keuls test showed this to be due to a significant difference in the amount of information recalled in the initial and readback phases (Q = 11.156 p < 0.01). Figures 11.1 and 11.2 give a visual indication of the difference in recall level during these interview phases. There was no significant difference between the amounts of information recalled in the initial and prompt phases (Q = 6.199 p > 0.05) or the prompt and readback phases (Q = 4.956 p > 0.05).

Another highly significant effect was the interaction between phase and information type recalled: F(10/420) = 21.3993 p = 0.0000 (see Appendix 32).

FIGURE 11.1 - VISUAL, CONVERSATIONAL AND FACTUAL INFORMATION x GROUP x PHASE

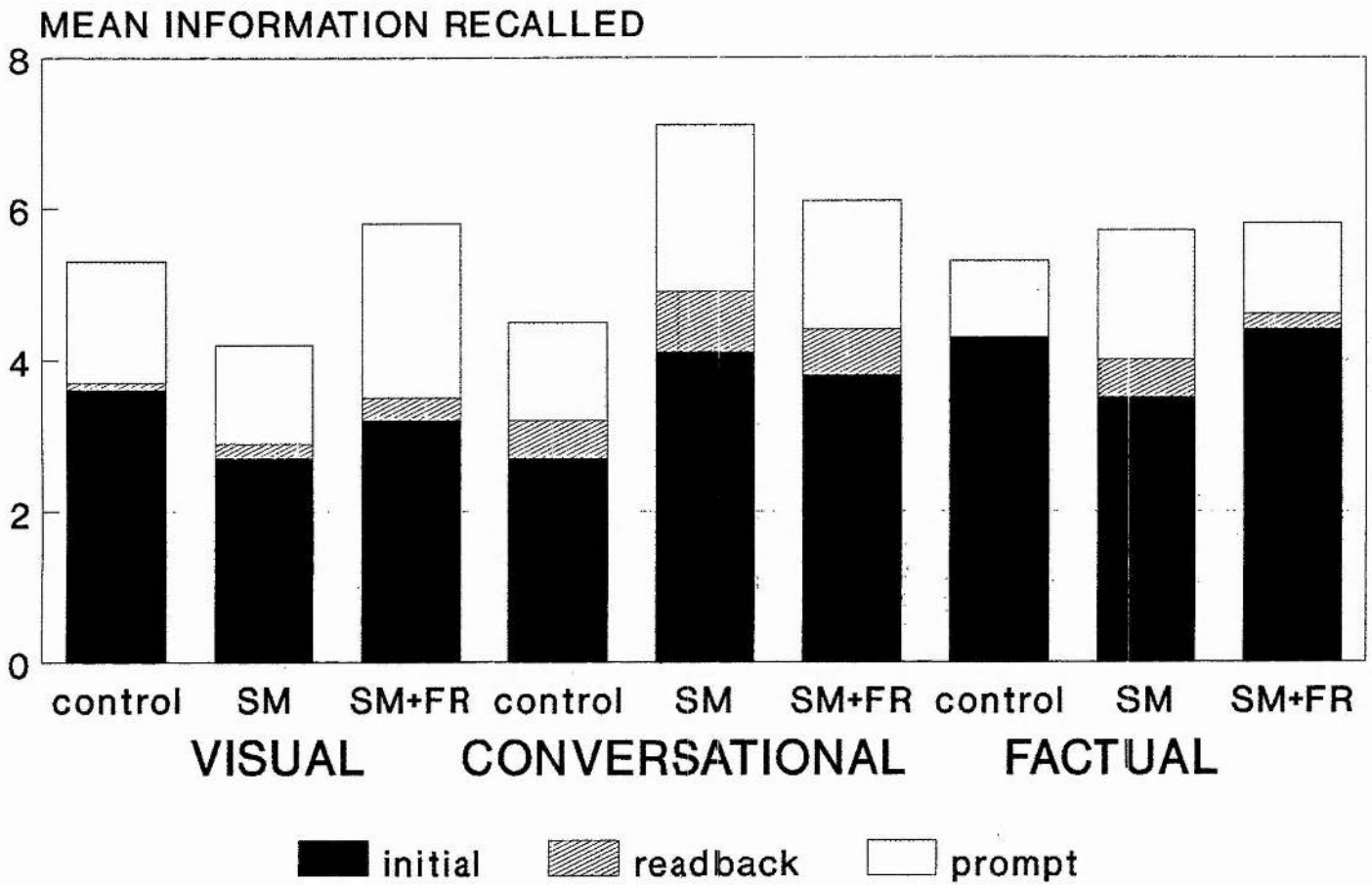
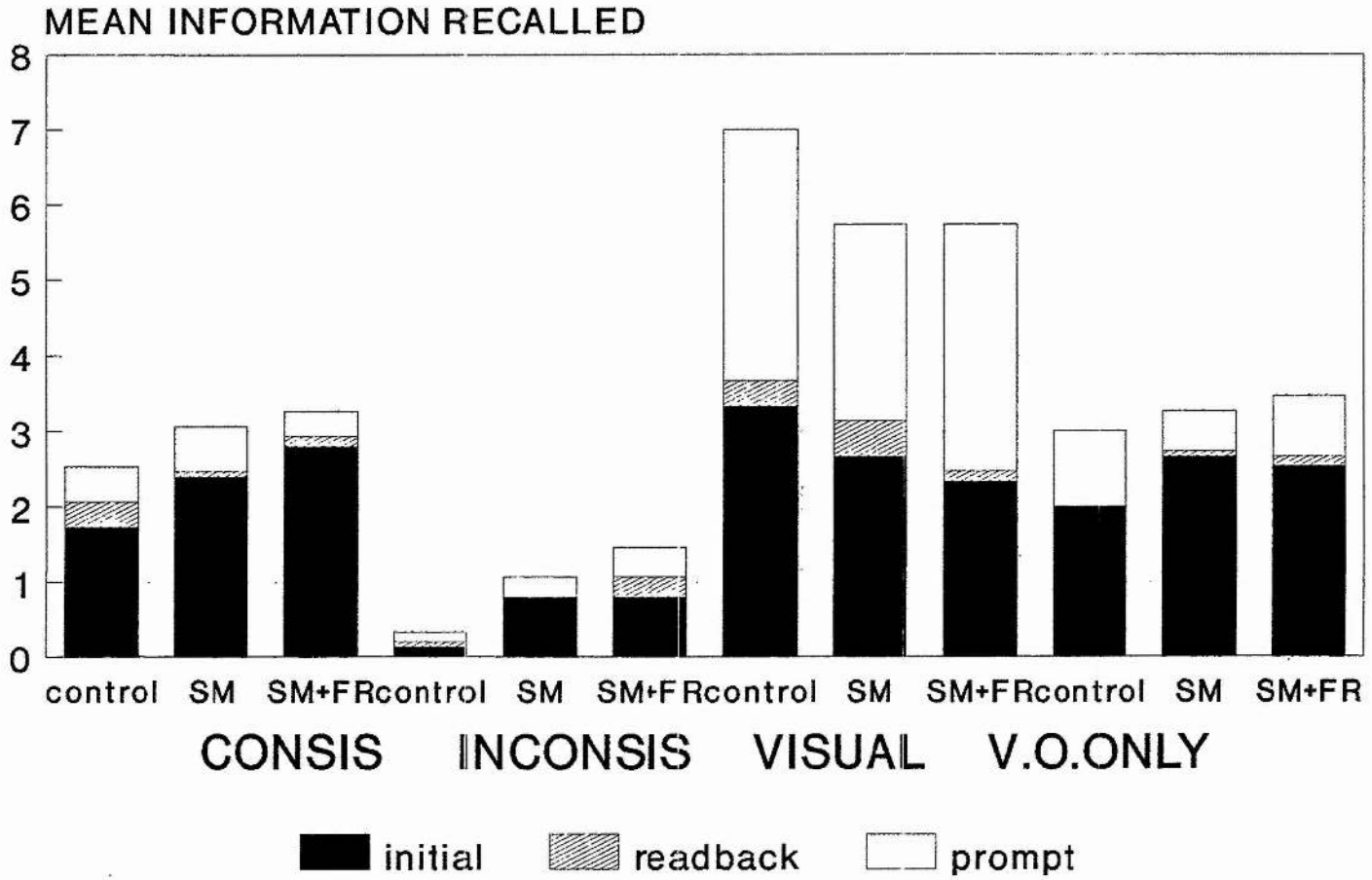


FIGURE 11.2 - FACTUAL INFO. RECALLED x INFORMATION TYPE x PHASE



A Neumann-Keuls test on these data showed that more conversational information was recalled in the initial phase than in the readback phase ($Q = 3.467$ $p < 0.01$), or the prompt phase ($Q = 2.333$ $p < 0.01$), also that, in the initial phase, more *conversational* information was recalled than any other type of information. More visual information was recalled in the prompt phase than the readback phase ($Q = 2.444$ $p < 0.01$); also more visual information was recalled in this phase than any other type of information. More visual information was recalled in the initial phase than in the readback phase ($Q = 2.2$ $p < 0.01$). More voice-over-only information was recalled in the initial phase than in the readback phase ($Q = 1.97$ $p < 0.01$) or the prompt phase ($Q = 1.578$ $p < 0.01$). More conversational information was recalled in the prompt phase than in the readback phase ($Q = 1.133$ $p < 0.01$ - see Figure 11.1 for a visual representation of these data).

Since there were no significant differences between each pair of means for the initial phase, analysis was carried out on the data generated in the final two phases only. Again, interview group was found to have no significant effect on information generated: $F(2/42) = 0.3457$ $p = 0.7098$. There was however a significant effect of phase: $F(1/42) = 140.2299$ $p = 0.0000$: more information was generated in the prompt phase than in the readback phase (the different levels of recall in these two phases is shown in Figures 11.1

and 11.2). There was no significant interaction between interview group and phase: $F(2/42) = 0.1886$ $p = 0.8288$, or between interview group and category of information generated: $F(10/210) = 0.8893$ $p = 0.5441$. There was however a significant main effect of information category: $F(5/210) = 55.9806$ $p = 0.0000$. This is unsurprising, since the amount of information available to be remembered varied with information category. Finally, a highly significant interaction was discovered between interview phase and information category: $F(5/210) = 31.0355$ $p = 0.0000$ (see Appendix 33). A Neumann-Keuls test showed that there was a significant difference between the amount of *visual* information generated in the prompt phase and the amount of other types of information generated in the prompt phase. There was also a significant difference between the amount of conversational information generated in the prompt phase and the readback phase ($Q = 1.289$ $p < 0.01$), and a significant difference between the amount of *conversational* information generated in the prompt and readback phases ($Q = 0.699$ $p < 0.01$ - see Figure 11.1).

Univariate analyses were then performed for each information type across phase and interview group (see Appendix 34). A significant effect of interview group was found for inconsistent information: $F(2/42) = 5.4772$ $p = 0.0077$. A Neumann-Keuls test was however unable to localise the source of this difference. Otherwise, the only significant effects found were for phase of interview. This was supported by the

initial, major analysis of variance, in which a highly significant effect of phase was discovered. Interview phase was found to have a significant effect on every type of information.

Highly significant effects for information type were found both in the initial analysis of variance: $F(5/210) = 87.0676$ $p = 0.0000$, and in the readback and prompt phases analysis of variance: $F(5/210) = 55.9806$ $p = 0.0000$. This was to be expected given the different amounts of information available across categories. Significant interactions between information type and phase were also found in the initial analysis of variance: $F(10/420) = 21.3993$ $p = 0.0000$, and in the readback and prompt phases analysis of variance: $F(5/210) = 31.0355$ $p = 0.0000$. Once again this was to be expected, given the differing performance levels in the three phases.

Errors:

Cond.		Initial	Readback	Prompt	Total
C	M	1.599	0.1333	3.133	4.867
	SD	0.9612	0.3519	2.987	3.378
SM	M	1.999	0.067	1.133	3.199
	SD	1.512	0.282	1.033	3.243
SMFR	M	0.933	0.133	1.600	2.667
	SD	1.254	0.414	1.056	3.570

TABLE 11.2 - MEAN ERRORS IN EACH INTERVIEW PHASE FOR EACH INTERVIEW TYPE, WITH STANDARD DEVIATIONS

Error data are also presented in Table 11.2, and also in histogram form in Figure 11.3 (showing errors made for different information types) and Figure 11.4 (showing different types of error made for factual information). An analysis of variance was performed on these data. It was found that interview group had no significant effect on the number of errors made: $F(2/42) = 1.2467$ $p = 0.2979$. Neither did phase of interview have a significant effect: $F(2/84) = 0.9519$ $p = 0.3901$. The interaction between phase and interview group was similarly non-significant: $F(4/84) = 2.3213$ $p = 0.0634$. The effect of information type was however highly significant: $F(5/210) = 24.5594$ $p = 0.0000$, although this was unsurprising since the amount of information available differed across information categories.

FIGURE 11.3 - VISUAL, CONVERSATIONAL AND FACTUAL ERRORS MADE x GROUP x PHASE

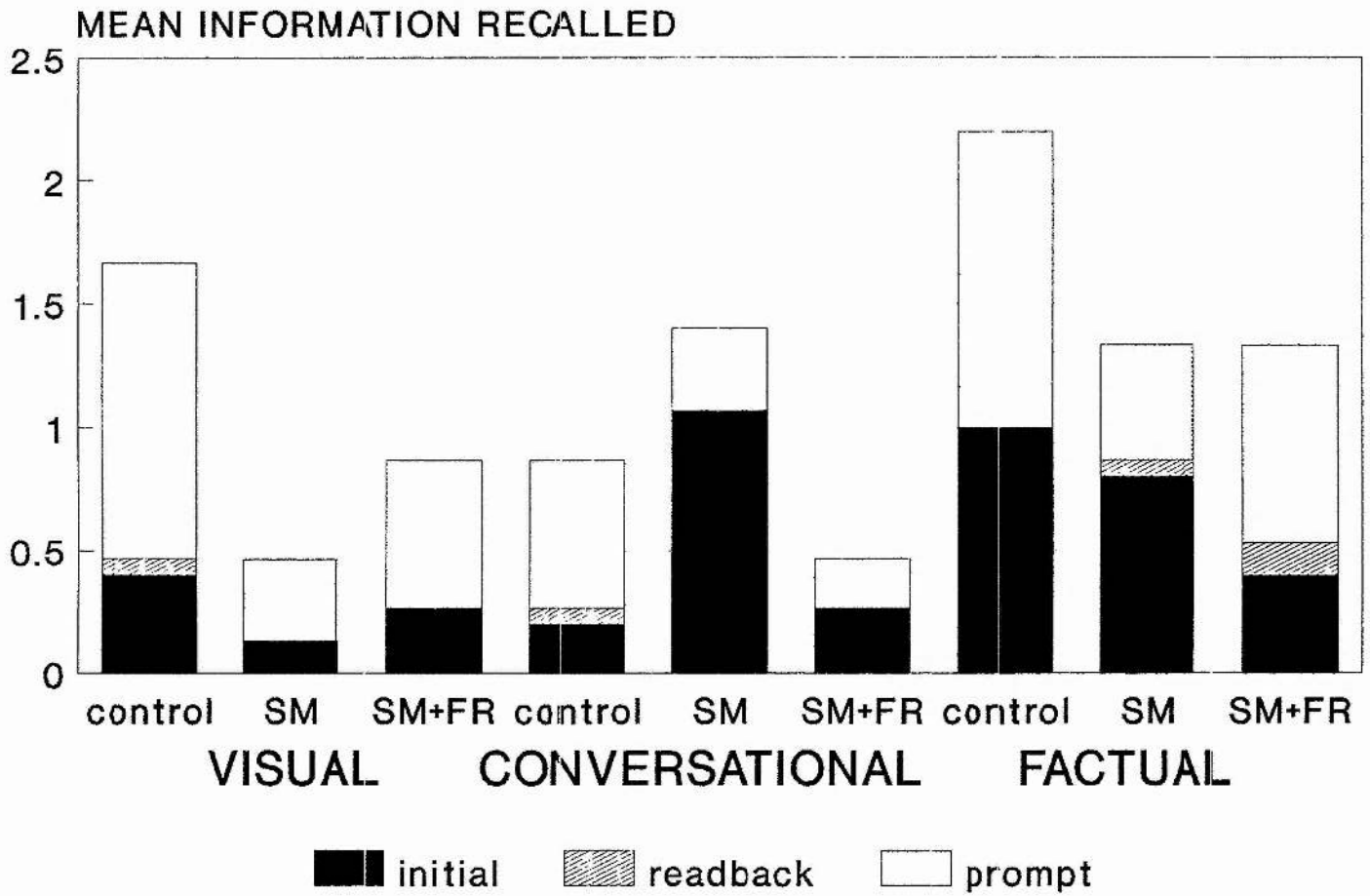
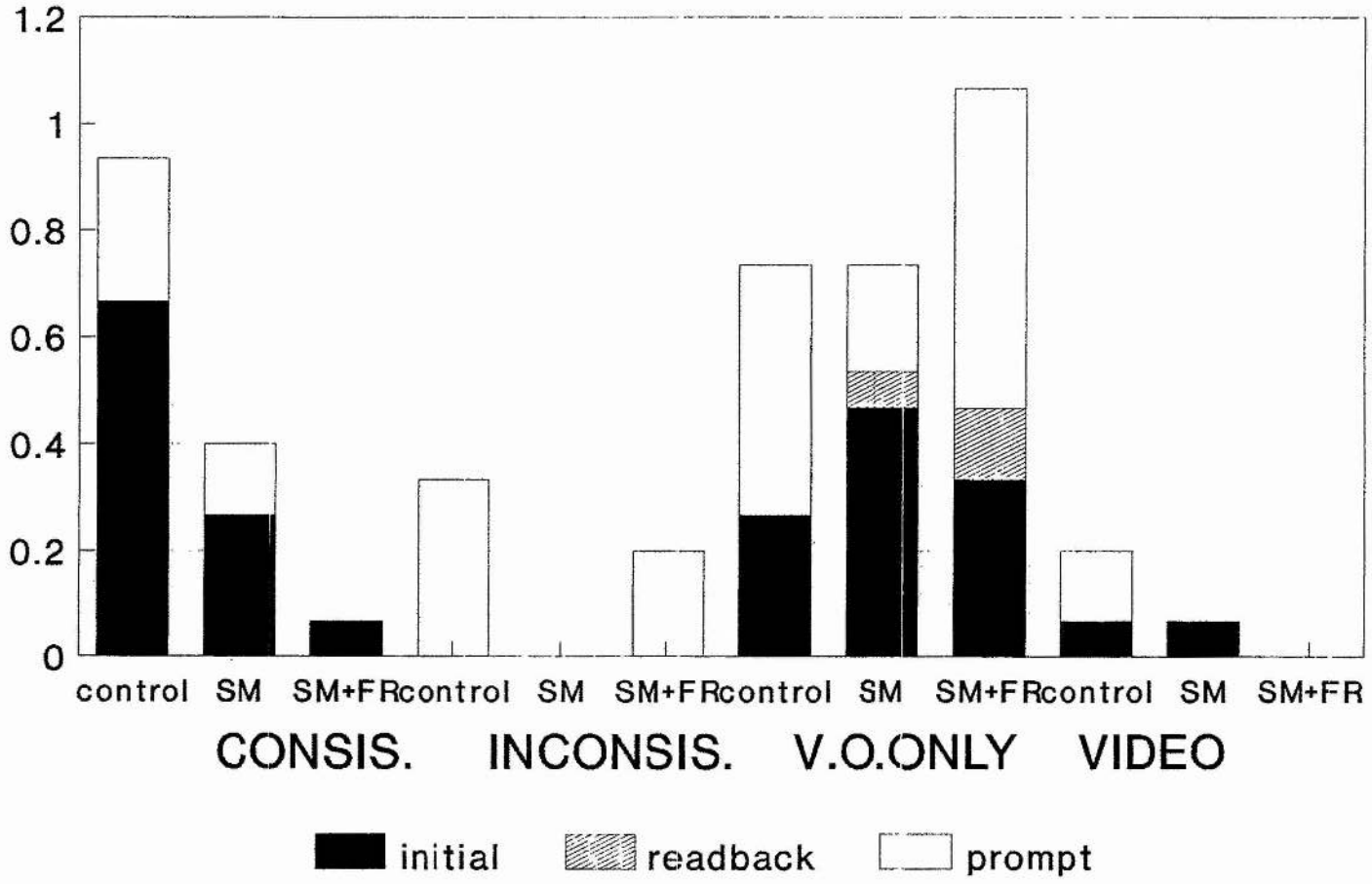


FIGURE 11.4 - FACTUAL ERRORS MADE x INFORMATION TYPE x PHASE



Also significant was the interaction between between interview group and information type: $F(10/210) = 3.4233$ $p = 0.0004$ (see Appendix 35). A Neumann-Keuls test revealed more visual errors to be made in Condition C than in either of the other conditions ($Q = 1.2$ $p < 0.01$ for Condition SM; $Q = 0.8$ $p < 0.01$ for Condition SMFR) and also that more visual errors were made than any other kind of error in Condition C.

More conversational errors were found to occur in Condition SM than in either of the other conditions ($Q = 0.533$ $p < 0.05$ for Condition C, $Q = 0.866$ $p < 0.01$ for Condition SMFR); also, more conversational errors than any other kind of error were made in Condition SM. These data are presented visually in Figure 11.3.

A significant interaction was found between information type and interview phase: $F(10/420) = 2.1998$ $p = 0.017$. A Neumann-Keuls test showed that more errors were made in the prompt phase for visual information than in either the initial or readback phases ($Q = 0.6889$, $p < 0.01$ $Q = 0.3556$ $p < 0.01$, respectively). Fewer conversational errors were made in the readback phase than in the initial phase ($Q = 0.4889$ $p < 0.01$) or the prompt phase ($Q = 0.3556$ $p < 0.05$). Equally, fewer voice-over-only errors were made in the readback phase than in the initial phase ($Q = 0.3111$ $p < 0.05$) or the prompt phase ($Q = 0.3556$ $p < 0.01$) - see Figure 11.3.

Finally, a significant interaction was found between interview group, information type, and phase of

interview: $F(20/420) = 1.8190$ $p = 0.0170$ (see Appendix 35). A Neumann-Keuls test on these data showed that more errors were made for visual information in the prompt phase of Condition C than in the prompt phase of Condition SM ($Q = 0.8667$ $p < 0.01$) or Condition SMFR ($Q = 0.6$ $p < 0.05$). There were also more errors made for this information type in Condition C than for any other information type in the prompt phase. More errors were made in the initial phase of Condition SM for conversational information than for any other type of information. More conversational errors were made in this phase of this condition than in the initial phase of Condition C ($Q = 0.8667$ $p < 0.01$) or of Condition SMFR ($Q = 0.8$ $p < 0.01$). No other significant differences were found.

Since an unexpected difference was found in the performance of subjects in the first phase for correct information, it was decided to perform an analysis of variance on errors made in the final two experimental phases only (see Appendix 36). A significant effect of interview group was found: $F(2/42) = 4.6148$ $p = 0.0154$. Significantly more errors were made in Condition C than in either of the other experimental conditions ($Q = 2.067$ $p < 0.01$ for Condition SM; $Q = 1.333$ $p = 0.01$ for Condition SMFR). No difference in the error rates in the experimental conditions was found. Once again, a significant effect was found for information type, although this was unsurprising since the number of bits of information available varied across categories: $F(5/210) = 18.5061$ $p = 0.0000$. One

other main effect was also significant, and that was the effect of interview phase: $F(1/42) = 10.2327$ $p = 0.0026$; more errors were made in the readback phase than in the prompt phase.

Two interactions were also significant: these were the interaction between interview group and information type: $F(10/210) = 2.2573$ $p = 0.0159$), and that between information type and phase of interview: $F(5/210) = 2.5681$ $p = 0.028$. Neumann-Keuls tests were performed to discover the loci of these two effects. For the interaction between interview group and information type, more visual errors were found to be made in Condition C than in Condition SM ($Q = 1.2$ $p < 0.01$) or Condition SMFR ($Q = 0.8$ $p < 0.01$). Within Condition C, more visual errors are made than any other type of error. More conversational errors were made in Condition SM than in Condition SMFR ($Q = 0.933$ $p < 0.01$), or Condition C ($Q = 0.533$ $p < 0.05$). As for the significant interaction between information type and phase of interview, it was found that significant differences lay between prompt and readback phases for visual information ($Q = 0.689$ $p < 0.01$) and also, within the prompt phase, between visual information and all the other information-types. More voice-over-only errors were made during the prompt phase than the readback phase ($Q = 0.3556$ $p < 0.01$); and also errors for conversational information ($Q = 0.3556$ $p < 0.01$).

Univariate analyses were performed for errors made for each information type, across interview groups and

phases (see Appendix 37). In the case of voice-over only information and inconsistent information, only the effect of phase was found to be significant, which was unsurprising given the differential performance of subjects across phases (see above). For video-only and consistent information, no significant effects were discovered. However, for conversational information, significant effects were found for interview group: $F(2/42) = 4.1774$ $p = 0.0221$; for phase: $F(2/84) = 8.1807$ $p = 0.0006$; and for the interaction between interview group and phase: $F(4/84) = 4.3223$ $p = 0.0031$. A Neumann-Keuls test on the effect of condition showed that significantly more errors for conversational information were made by subjects in Condition SM than by subjects in Condition C ($Q = 0.5 \cdot 33 \cdot$ $p < 0.05$) or by subjects in Condition SMFR ($Q = 0.9 \cdot 33 \cdot$ $p < 0.05$). More errors were made by subjects in Condition C than by subjects in Condition SMFR ($Q = 0.4$ $p < 0.05$). A Neumann-Keuls test on the interaction between interview group and phase showed that significantly more errors were made by subjects in the initial phase of Condition SM than by subjects in any other phase of any other condition. There were no other significant effects.

Accuracy:

Accuracy was calculated as:

$$\frac{\text{Total amount of correct information}}{\text{Total amount recalled}}$$

The table below gives mean percentage accuracy for each phase for the three interview groups:

	Initial	Readback	Prompt	Total
C	89.879	91.667*	76.020	85.856
SM	90.881	96.296**	84.259	90.479
SMFR	93.805	92.263***	85.799	90.622

*N = 9 **N = 13 ***N = 10. In each readback phase, a number of subjects recalled no correct information and made no errors. Since accuracy scores for these subjects could not be computed, their scores were omitted from the analysis.

TABLE 11.3 - MEAN ACCURACY OF RECALL x INTERVIEW GROUP x INTERVIEW PHASE

An analysis of variance was performed on these data. Since there were different numbers of subjects in each readback phase, phase could not be considered as a *within*-subjects variable. Hence the data was analysed as if both interview group *and* phase were *between*-subjects factors. There was no significant effect of interview group: $F(2/113) = 1.3949$ $p = 0.2521$, and no significant interaction of interview group with phase: $F(4/113) = 0.6444$ $p = 0.6320$. There was however a main effect of interview phase: $F(2/113) = 6.7977$ $p = 0.0016$ (Appendix 38). A Neumann-Keuls test showed the difference to lie between the prompt phase and the two other phases. Between the prompt phase and the readback phase, $Q = 13.962$ $p < 0.01$; between the prompt phase and the initial phase, $Q = 12.507$ $p < 0.01$. Accuracy of recall in the prompt phase was hence significantly lower than in either of the other two phases.

In this case, it is more appropriate to discount those subjects who recalled no information and made no

errors during the readback phase of the interviews; since their contribution during this phase would be of no interest to a police officer investigating a crime. However, it is possible to hold that these subjects, having made no errors during the phase, were in fact completely accurate. If their accuracy scores during this phase are counted as 100%, then a different table of mean values is produced. In Table 11.4 below, the mean accuracy values for the readback phase are rather higher than those given in Table 11.3.

	Initial	Readback	Prompt	Total
C	89.879	95.000	76.020	85.856
SM	90.881	96.790	84.259	90.479
SM+FR	93.805	94.842	85.799	90.622

TABLE 11.4 - MEAN ACCURACY OF RECALL x INTERVIEW GROUP x INTERVIEW PHASE, INCLUDING PREVIOUSLY DISCOUNTED DATA

Monitoring the Source of Information: Source-monitoring took place during either the readback phase (in Condition SM) or the initial phase (in Condition SMFR). During their respective source-monitoring phases, subjects in Conditions SM and SMFR were asked to attribute sources for the items of factual information that they remembered. The number of correct source allocations made was divided by the total amount of correct information given in that category by each subject, to give a percentage accuracy score. These are shown in Table 11.5:

	No. of subjects giving info. in the category	Mean accuracy %
Condition SM		
VO Only	15	91.11
Consistent	14	80.67
Inconsistent	8	87.50
Video Only	8	62.50
Condition SMFR		
VO Only	14	93.45
Consistent	14	100.00
Inconsistent	9	96.30
Video Only	5	100.00
Total	30	88.94

TABLE 11.5 - MEAN ACCURACY SCORES FOR EACH TYPE OF FACTUAL INFORMATION x CONDITION, PLUS NUMBER OF SUBJECTS GIVING INFORMATION IN EACH CATEGORY

Hence, total mean accuracy across conditions and items of factual infomation was 88.94%. A single-sample t-test showed this to be significantly above the chance score of 33.33%, $t(29) = 7.39$ $p < 0.001$.

The Effect of Stereotypes on Recall: Of forty-five subjects, thirty-five correctly identified the profession of one of the characters in the video-recording as a lawyer, and thirty correctly identified the profession of one of the characters as a labourer. Of the remainder, eight subjects misremembered another white-collar profession in place of "lawyer", and only two remembered no detail about that profession. In the case of the labourer, two subjects misremembered another blue-collar profession, and three substitutions of different occupations occurred (specifically, one subject remembered one character as being a disc-jockey, another "an advisor or something-

or-other", and one as being "on the dole"). Ten subjects remembered nothing about this character's profession.

Of those who identified one of the characters as a lawyer, 100% (i.e. thirty-five subjects) thought that he was the man in the suit. Similarly, of those who identified one of the characters as a labourer, 100% (i.e., thirty) thought that he was the casually-dressed man.

Discussion

There follows a brief discussion of each of the experimental hypotheses in turn:

1. The first hypothesis was confirmed insofar as more pieces of correct information were recalled by subjects in Condition SMFR than by those in Condition C. However it should be borne in mind that, during the initial phase, more pieces of correct information were recalled by subjects in Condition SM than by those in Condition SMFR: i.e., before any manipulation had taken place. As far as errors are concerned, no effect was found during the initial phase for any interview type. This was also the case for accuracy of recall during this phase. Hence it must be concluded that this hypothesis was not, overall, confirmed.

Interestingly, very similar results were reported by Sussman (1992), who found that subjects in the source-monitoring condition recalled significantly more pieces of correct information than subjects in the

control condition during the initial phase, despite the fact that no manipulation had yet taken place; also that more errors were made by these subjects in this phase, and that mean percentage accuracy appeared to be lower (although a Neumann-Keuls test could not localise the difference discovered here through analysis of variance).

2. Subjects in Condition SM were not found to recall more correct information than those in Condition C during the readback phase, and, indeed, there was found to be a trend towards more correct information being recalled in Condition SMFR during this phase than in Condition SM. Neither were fewer errors made during this phase for subjects in Condition SM, nor was accuracy significantly greater in this phase for subjects in this condtion.

Similarly, Sussman (1992) found no significant difference between the two groups in the readback phase for correct information, for errors made, or for accuracy of recall.

3. The use of the source-monitoring technique did not aid memory for factual information; equal numbers of errors were made for factual information in the two source-monitoring groups as in the control group; the only effects found were for conversational and for visual information. Again, this roughly parallels the results of Sussman (1992), who found a significant effect only for conversational information.

Of the subjects who were able to remember either one or both of the professions of the characters in the

video-recording, 100% made the stereotype-consistent error of saying that the man in the suit was the lawyer and/or that the casually-dressed man was the labourer. Hence, no effect of source-monitoring instructions was discovered here. Of these, subjects who misremembered the professions specified, only one substituted a white-collar job ("an advisor or something-or-other") for the blue-collar job of labourer.

4. Analysis of scores for accuracy of monitoring the source of information shows that subjects were able to perform the source-monitoring task with a significant degree of accuracy: 88.941% accuracy in monitoring the sources of information being a significantly higher score than would be expected by chance. Hence any failure by subjects in these conditions to fulfill other hypotheses predicted for them cannot be explained by the cognitive complexity of the task: subjects can and do carry out the task when asked.

5. Although the trend was for subjects in the source-monitoring conditions to recall more pieces of correct information than those in the control condition (4.2 pieces/subject for the source-monitoring task, 3.933 pieces for the source-monitoring plus free recall task) this difference was not found to be significant. Between Conditions SM and SMFR, the trend was in the opposite direction from that predicted by the hypothesis, although the difference in scores was minimal (in fact, 0.267 of one piece of information). As far as number of errors was concerned, the trend

was in the predicted direction: Condition C subjects made a mean 0.8 more errors than Condition SM subjects, who in turn made a mean 0.2 more errors than Condition SMFR subjects. This trend was not however found to be significant. Subjects in Condition SM were a mean 7.551% more accurate than control subjects - a trend predicted by the experimental hypotheses - but also a mean 0.55% more accurate than subjects in Condition SMFR, this trend being in the opposite direction from that predicted. Neither of these differences was however significant.

Hence it must be concluded that the present experiment gives little support to the notion that a source-monitoring or a source-monitoring plus free-recall interview could be used to improve an eyewitness' recall performance for information that has been distorted by stereotypic information. A small amount of evidence was collected in favour of the hypotheses suggested by such a theory; however only a few of the predicted effects attained levels of significance.

This result does help to round out the emerging theory of stereotypes and eyewitness memory. Three earlier experiments in the present thesis show the effects of stereotypes on memory when testing occurs immediately after presentation of to-be-remembered material (to wit, Experiments 2 and 3, and Experiment 4): and these were the *only* three to show such effects. The two experiments in which a time-lag was involved between stimulus presentation and testing

showed no effect of stereotypes on memory (to wit, Experiment 6, in which delays of six months and one year were used, and Experiment 5, in which a delay of three weeks was used).

It appears likely that where stereotypes have an effect on the memory trace (this in turn appears more likely to occur when stereotypic information is presented at encoding than when it is presented at retrieval - see Experiments 4 and 5, and also Chapter 6), this effect will decay with time, being most robust at the point immediately following presentation of the material. Therefore, interviews will more likely be efficacious when conducted after a delay than when conducted immediately after witnessing an event, in terms of minimising potential biassing effects of stereotypic information. Given the amount of support the source-monitoring interview has gained in the literature, and the number of successful studies that have been carried out employing the technique (e.g., Johnson *et al*, 1979, 1981; Foley and a Johnson, 1982; and, especially, Lindsay and Johnson, 1989a, and b; also Lindsay and Johnson, 1989c, cited in Lindsay and Johnson, 1989a) this promises to be the most fruitful next avenue to explore.

CHAPTER 12 - CITING YOUR SOURCES: USE OF THE SOURCE-MONITORING TECHNIQUE IN A DELAYED-TEST CONDITION

Experiment 8

Introduction

Experiment 7, above, considered the efficacy of source-monitoring and source-monitoring-plus-free-recall interviews as means of reducing the impact of stereotypic information on memory for a social event. Although the hypotheses laid down with regard to these interviews were not on the whole fulfilled, the possibility was mooted that the interviews might more likely prove effective when there was a delay between presentation of to-be-remembered material and the interview. Results of previous experiments imply that such a delay may allow the degradation of biassing or distorting effect of stereotypes on social memory: it is under these conditions that a source-monitoring interview may access the original information only. Should this effect hold, it will imply that any model that might be proposed of the effect of stereotypes on social memory must take account of the temporal dimension.

This experiment has a second aim. Experiment 7 aimed to generate data to test the efficacy of the source-monitoring interview in reducing the impact of stereotypic information on memory for social information. However, the present experiment can also be looked upon as a test of a proposal put forward by Wright (1992) in explanation of the results obtained in studies of that ongoing area of controversy in

eyewitness research, the misinformation paradigm. Wright argues cogently that:

> "...when the different items have different sources that are not contradictory, they may harmoniously coexist, but if the source traces vanish the items will contradict and a blend may need to occur" [*sic*] (Wright, 1992, p 20)

In a classic Loftus experiment, therefore, in which, e.g., a "yield" sign is presented in the slide sequence and a "stop" sign used for misinformation in a latterly-presented prose account, Wright holds that while the memory of the two information sources (slide sequence and prose account) is still present in the subjects' minds then both pieces of information are also present and either can be accessed. After time, however, when the memory sources have faded, a blend occurs and neither piece of information (i.e., in this case, "yield" or "stop" sign) can be accessed independently. This claim is backed up by a number of experimental observations, e.g., Lindsay (1990) who used a two-day lag between stimulus presentation and testing and found that the original information could not now be accessed. A comparison of this result with that obtained by Christiaansen and Ochalek (1983) is informative. Both Lindsay, and Christiaansen and Ochalek, investigate the efficacy of *warnings* in reducing the misinformation effect (i.e., specifically telling subjects that they have been presented with misinformation). Christiaansen and Ochalek find that a warning presented immediately after presentation of misinformation is effective in reducing the misinformation effect; after two days however a

warning appears to be less effective (Lindsay, 1990). In the context of police interviewing, Belli *et al* (1982) write: "The temporal dimension stresses the importance of interrogating witnesses soon after the event".

Wright himself tests his model in a series of experiments in which slide sequences are used for stimuli, and the model appears to hold up very well.

Although the present experiment does not conform to the classic misinformation paradigm, it is nevertheless conceptually close enough to allow inferences to be drawn from it to the misinformation situation, and vice versa (see Chapers 6 and 11, above). Subjects' stereotypes of the personal appearance of lawyers and labourers can be considered the original information, and the information given on the voice-over as to their actual professions can be considered the misinformation. If Wright's hypothesis is correct, then after the delay a blend should occur between the "original information" and the "misinformation". Since the occurrence of a blend depends upon loss of information about the sources of the two types of information, then source-monitoring should, *ex hypothesi*, have no effect on memory for the information.

In summary therefore, if the Wright model holds, then in an experiment which conceptually replicates Experiment 7, except for the inclusion of a delay between the presentation of target material and testing, then no group of subjects should differ

significantly in its results from any other. However, if the Wright model does *not* hold, and the different sources are still discriminable by subjects after a delay, then subjects in the source-monitoring-plus-free-recall condition (Condition SMFR) are expected to perform significantly better (i.e., producing more pieces of correct information, making fewer errors, and having greater accuracy,) than subjects in the source-monitoring condition (Condition SM), who, in turn, are expected to perform significantly better than subjects in the control condition (Condition C).

Otherwise, the hypotheses of the present experiment are:

1. In the initial phase, subjects in Condition SMFR should recall more information, make fewer errors, and be more accurate, than subjects in either of the two other conditions.

2. In the readback phase, subjects in Condition SM should remember more information, make fewer errors, and be more accurate, than subjects in Condition C.

3. Subjects in Conditions SM and SMFR should make fewer errors for factual information than those in either of the other two conditions, and make fewer stereotype-consistent errors regarding the true professions of the characters featured on the video-recording.

4. Overall, performance in Conditions SM and SMFR should be characterised by greater recall of correct information, greater accuracy, and fewer errors than performance in Condition C; recall performace in

Condition SMFR should be similarly superior to performance in Condition SM.

Methodology

Subjects: Thirty students (sixteen female, fourteen male), both undergraduates and postgraduates, aged 18 - 27 years, participated in the study for the chance of winning first or second prize in a raffle which was limited to the set of experimental subjects.

Materials: These were identical with those used in the previous study.

Procedure: Subjects participated in this experiment one at a time. The procedure was identical with that for the previous experiment, except that stimulus exposure and recall were separated by a period of three weeks. All interviews were tape-recorded for later analysis, and no subject was de-briefed until the end of the second experimental session.

Scoring Procedure: This was identical with that used in Experiment 7.

Results

Correct Information:

The raw data for correct information generated by interview group and interview phase are given in Table 12.1. A breakdown of this data by information type is

given in Figure 12.1; and a breakdown of the types of factual information generated in the four interview phases is given in Figure 12.2.

Cond.		Initial	Readback	Prompt	Total
	M	10.600	0.600	3.900	15.100
C	SD	2.503	1.732	1.792	4.198
	M	12.400	1.500	5.200	19.100
SM	SD	4.584	1.345	2.440	5.306
	M	11.200	1.100	5.000	17.300
SMFR	SD	3.343	1.134	1.826	4.606

TABLE 12.1 - MEAN CORRECT INFORMATION REMEMBERED IN EACH INTERVIEW PHASE FOR EACH INTERVIEW TYPE, WITH STANDARD DEVIATIONS

The same process of analysis was applied to these data as to those in Experiment 7. First, analysis of variance was performed on the entire data set. There was no significant effect of interview group on the amount of information remembered: $F(2/27) = 1.7709$ $p = 0.1894$; this was the same result as found in Experiment 7. This time however there was no significant interaction between interview group and interview phase for the amount of information remembered: $F(4/54) = 0.2106$ $p = 0.9315$, in contradiction to the results of Experiment 7. Hence, no differential effect of the positioning of source-monitoring instructions (whether in the initial or readback phase) on subjects' performance in that phase was found.

FIGURE 12.1 - VISUAL, CONVERSATIONAL & FACTUAL INFORMATION x GROUP x PHASE

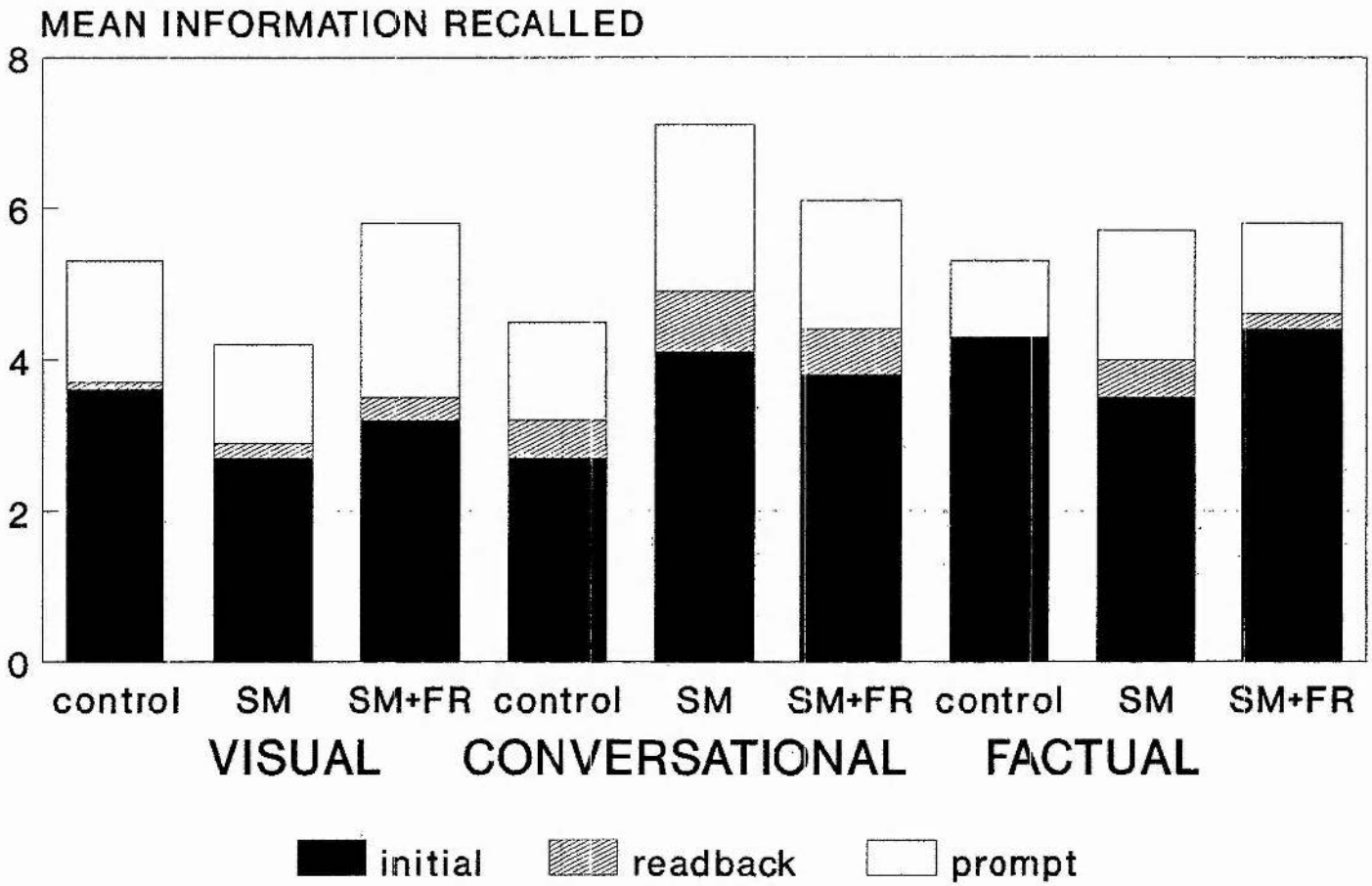
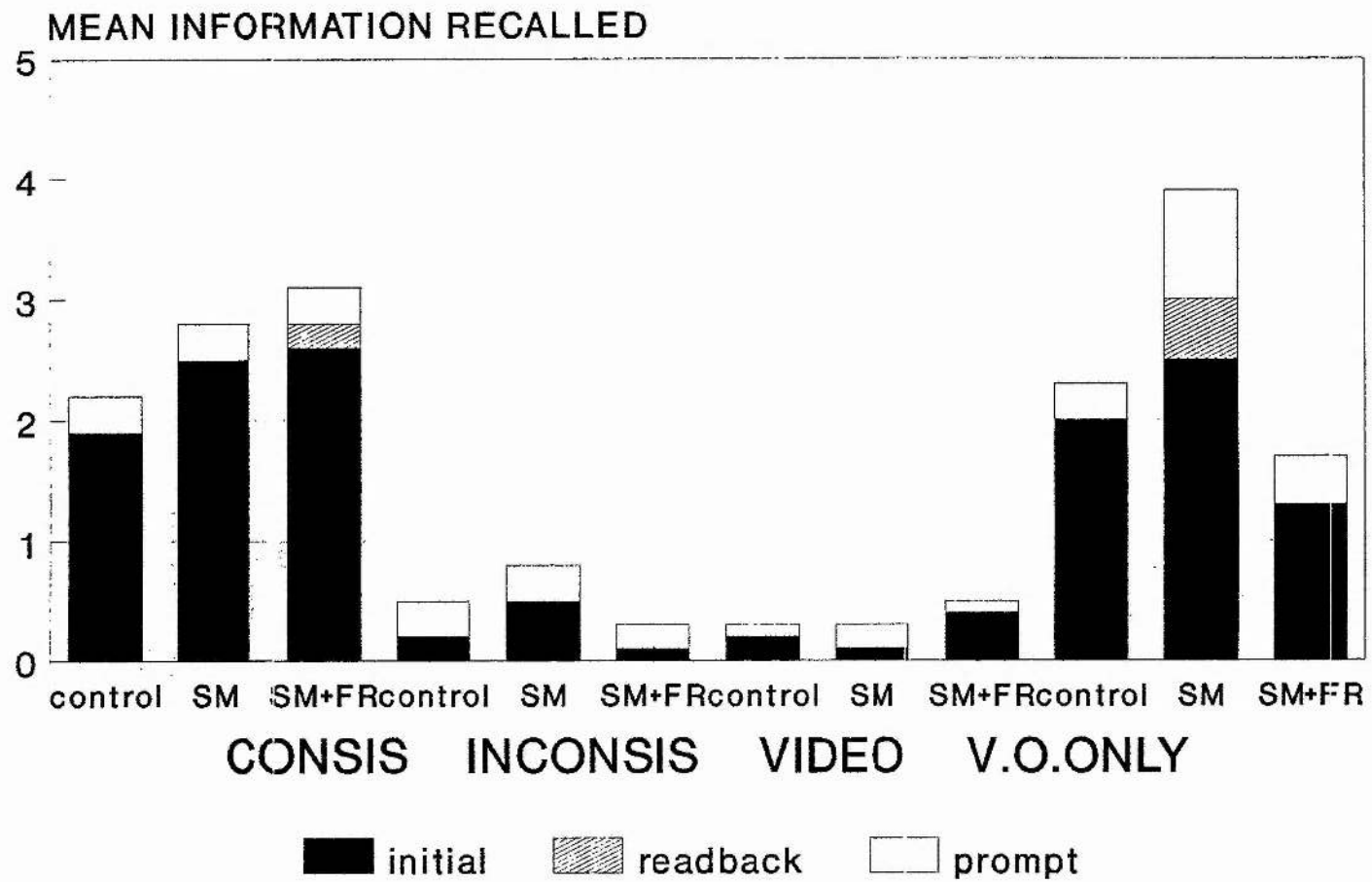


FIGURE 12.2 - FACTUAL INFO. RECALLED x INFORMATION TYPE x PHASE



Since no significant differences were found between the control and the source-monitoring groups in the initial phase, there was no necessity in this experiment to perform separate analysis of the results obtained in the final two phases only, as in Experiment 7.

There was a significant interaction between interview group and type of information remembered: $F(10/135) = 2.5154$ $p = 0.0083$ (see Appendix 39). However, a Neumann-Keuls test was unable to isolate the source of this interaction: no one interview type was found to produce significant gains or losses for any specific type of information over any other interview type (see Figures 12.1 and 12.2).

Again, a highly significant effect of phase was found: $F(2/54) = 177.5376$ $p = 0.0000$. A Neumann-Keuls test showed this to be due to a significant difference between the amount of correct information recalled in every phase compared with every other phase. More correct information was recalled in the initial phase than in the readback phase ($Q = 10.367$ $p < 0.01$) or the prompt phase ($Q = 6.7$ $p < 0.01$). More correct information was recalled in the prompt phase than in the readback phase ($Q = 3.667$ $p < 0.01$). Again, this is visually presented in Figures 12.1 and 12.2.

Again, a highly significant interaction was discovered between phase of interview and type of information recalled: $F(10/270) = 16.2166$ $p = 0.0000$. A Neumann-Keuls test showed again that more conversational information was recalled in the initial

phase than in the readback phase ($Q = 2.6$ $p < 0.01$) or the prompt phase ($Q = 1.8$ $p < 0.01$). More conversational information was recalled in the initial phase than was any other type of information: except for visual information ($Q = 0.366$ $p > 0.05$). More visual information was recalled in the initial phase than in the readback phase ($Q = 2.967$ $p < 0.01$) or the prompt phase ($Q = 1.434$ $p < 0.01$). More visual information was recalled in the initial phase than any other type of information (except conversational information). More consistent information was recalled in the initial phase than in the readback phase ($Q = 2.2$ $p < 0.01$) or the prompt phase ($Q = 1.434$ $p < 0.01$). More voice-over-only information was recalled in the initial phase than in the readback phase ($Q = 1.766$ $p < 0.01$) or in the prompt phase ($Q = 1.433$ $p < 0.01$). More visual information and conversational information was recalled in the prompt phase than in the readback phase ($Q = 1.533$ $p < 0.01$; $Q = 1.1$ $p < 0.01$, respectively). The number of bits of visual information recalled in the prompt phase did not differ significantly from the number of bits of conversational information recalled in that phase; however, the number of bits of *each type* of information did differ significantly from the number of bits of any of the other four types of information recalled in that phase.

As in Experiment 7, a highly significant effect of information type was discovered: $F(5/135) = 59.6439$ $p = 0.0000$. This is however unsurprising since the

amount of information available differed across categories.

Univariate analyses were then performed for each information type by interview group and phase of interview (see Appendix 40). In most cases, only phase of interview was found to give a significant effect: unsurprisingly, given that little information was recalled in the readback phase (see Table 12.1). For voice-over-only information, however, significant effects were found for both phase and interview group. A Neumann-Keuls test showed that subjects in Condition SM recalled significantly more correct voice-over-only information than subjects in Condition C ($Q = 1.6$ $p < 0.01$) or Condition SMFR ($Q = 2.2$ $p < 0.01$).

Errors:

The raw data by interview group and interview phase are given in Table 12.2, below. A breakdown of errors made by information type is given in Figure 12.3, and a breakdown of errors made for each factual information type is given in Figure 12.4.

Analysis of variance was performed on the whole data set. Interview group was found to have a significant effect on the number of errors made: $F(2/27) = 9.2840$ $p = 0.0009$ (Appendix 41). A Neumann-Keuls test showed that significantly more errors were made in Condition C than in the Condition SM ($Q = 3.4$ $p < 0.01$) or Condition SMFR ($Q = 4.7$ $p < 0.01$).

FIGURE 12.3 - VISUAL, CONVERSATIONAL AND FACTUAL ERRORS MADE x GROUP x PHAS

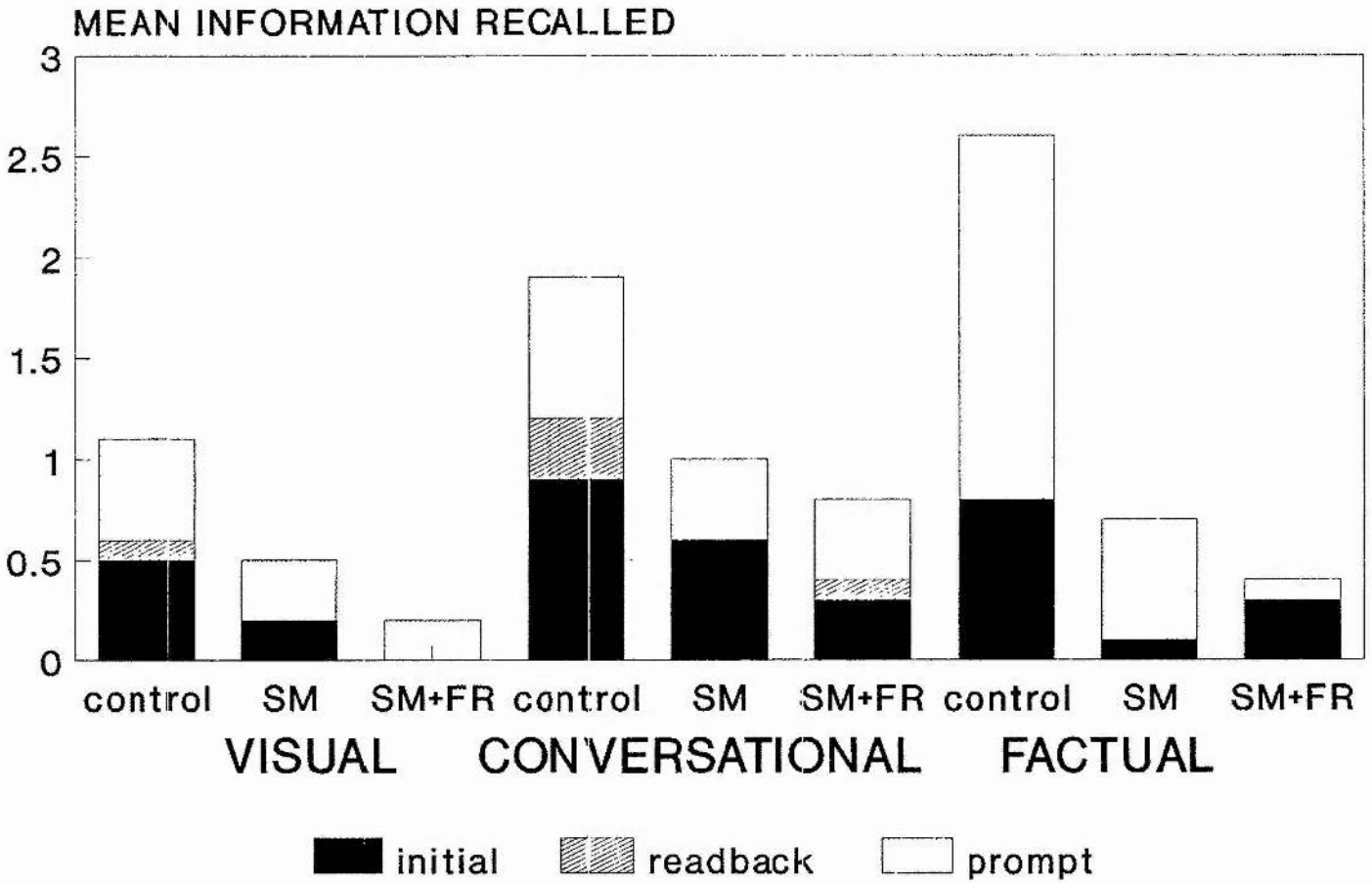
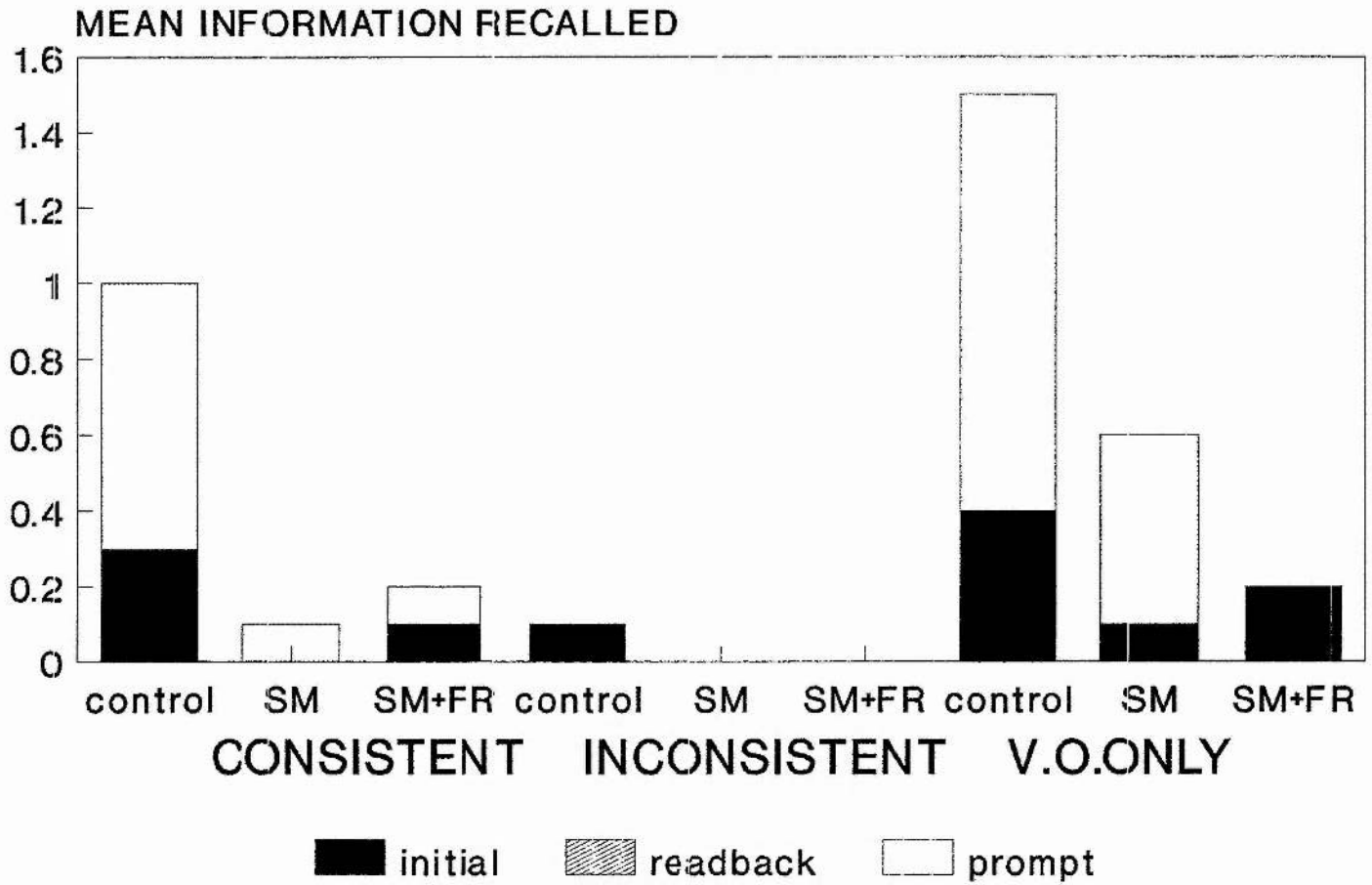


FIGURE 12.4 - FACTUAL ERRORS MADE x INFORMATION TYPE x PHASE



Additionally, more errors were found to have been made in Condition SM than in Condition SMFR ($Q = 1.3$ $p < 0.01$).

Cond.		Initial	Readback	Prompt	Total
	M	2.2	0.4	3.1	5.7
C	SD	2.099	0.577	1.590	1.794
	M	0.9	0.1	1.3	2.3
SM	SD	0.837	0.000	0.675	0.743
	M	0.6	0.1	0.7	1.4
SMFR	SD	0.447	0.000	1.140	0.527

TABLE 12.2 - MEAN ERRORS IN EACH INTERVIEW PHASE FOR EACH INTERVIEW TYPE, WITH STANDARD DEVIATIONS

There was no significant interaction between interview group and phase of interview, however: $F(4/54) = 2.2959$ $p = 0.0709$. Hence, number of errors made in each phase was unaffected by the position of the source-monitoring task (whether in the initial phase, the readback phase, or not present at all).

A significant effect of interview phase was also discovered: $F(2/54) = 14.5889$ $p = 0.0000$, in contradiction to Experiment 7. A Neumann-Keuls test showed that significantly more errors were made in the prompt phase than either the initial ($Q = 1.5$ $p < 0.01$) or readback phases ($Q = 1.033$ $p < 0.01$). There was a highly significant effect of information type: $F(5/135) = 10.6058$ $p = 0.0000$, but this was unsurprising given the different amounts of information available in each information category. There was no significant interaction between interview group and information type: $F(10/135) = 1.4051$ $p = 0.1844$; but there was a significant interaction

between interview *phase* and information type: $F(10/270) = 3.3915$ $p = 0.0003$ (see Figures 12.3 and 12.4). A Neumann-Keuls test on these data was however unable to localise the source of this effect.

There was no significant three-way interaction between group, phase, and information type: $F(20/270) = 1.5285$ $p = 0.0713$.

Univariate analyses were then performed for errors made for each information type, across phase and interview condition (Appendix 42). Visual information was not considered, since no errors were made for this information type in any condition. For conversational and visual information, only the effect of phase was significant (these data are presented visually in Figure 12.3). For voice-over-only information however there was a significant effect of interview group: $F(2/27) = 8.2552$ $p = 0.0016$; of interview phase: $F(2/54) = 12.2324$ $p = 0.0000$; and interaction between interview group and phase: $F(4/54) = 5.1021$ $p = 0.0015$ (see Figure 12.4). A Neumann-Keuls test showed that more errors were made by subjects in Condition C than by subjects in Condition SM ($Q = 0.9$ $p < 0.01$), or Condition SMFR ($Q = 1.3$ $p < 0.01$). It also revealed that more errors were made for voice-over-only information by subjects in Condition SM than subjects in Condition SMFR ($Q = 0.4$ $p < 0.05$) - see Figure 12.4.

The same pattern of effects held for stereotype-consistent information: there was a significant effect of interview group: $F(2/27) = 7.7294$ $p = 0.0022$; of

phase: $F(2/54) = 6.0330$ $p = 0.0043$; and a significant interaction between the two: $F(4/54) = 2.7692$ $p = 0.0363$. A Neumann-Keuls test showed that more errors were made for consistent information in Condition C than in Condition SM ($Q = 0.9$ $p < 0.01$) or Condition SMFR ($Q = 0.8$ $p < 0.01$). For the interaction, it revealed that more consistent errors were made in the prompt phase by subjects in Condition C than in either of the other phases or interview conditions (see Figure 12.4). There was no other significant effect. Hence, it appears that the effect of interview group on number of errors made is attributable at least in part to a reduction in the number of errors made for voice-over-only information across groups, and a reduction in number of errors made for consistent information with each source-monitoring interview. These significant effects are probably not the entire story however: a contributing factor to the significant effect discovered for interview group in the major analysis of variance (above) was also probably the general direction of the results for the other information types: errors decreasing from Condition C to Condition SMFR.

Accuracy:

Mean percentage accuracy for each phase for the three interview groups is given below. In Table 12.3, data for those subjects who produced no correct information but made no errors are discounted.

	Initial	Readback	Prompt	Total
C	85.456	46.667*	57.331	60.151
SM	93.083	100.000**	77.484	90.189
SMFR	95.065	97.143***	90.773	94.327

*N = 5, **N = 7, ***N = 7.

TABLE 12.3 - MEAN ACCURACY OF RECALL x INTERVIEW GROUP x INTERVIEW PHASE

Analysis of variance was performed on these data, with, once again, phase considered as a between-subjects variable (see Appendix 43). A significant effect of interview group on recall accuracy was discovered: $F(2/70) = 24.6202$ $p = 0.0000$. A Neumann-Keuls test showed that source-monitoring interviews elicited more accurate recall than the control interview, ($Q = 27.566$ $p < 0.01$ for Condition SMFR, $Q = 22.6508$ $p < 0.01$ for Condition SM,) but that the two source-monitoring interviews did not differ significantly from each other in terms of accuracy of information elicited ($Q = 4.9150$ $p > 0.05$). This contrasts with the results of Experiment 7, in which no significant effect of interview group was found.

A significant effect of interview phase was also discovered: $F(2/70) = 5.6067$ $p = 0.0055$. A Neumann-Keuls test showed that there was greater accuracy of recall in the initial than the prompt phase: $Q = 16.006$ $p < 0.01$. There was no other significant difference. This result is somewhat similar to that found in the previous experiment, in which accuracy of recall was found to be significantly lower in the prompt phase than in either of the other two phases.

Finally, a significant interaction was discovered between interview group and interview phase: $F(4/70) = 4.9995$ $p = 0.0013$. A Neumann-Keuls test showed accuracy to be lower in the readback phase of Condition C than in the readback phase of either Condition SM: $Q = 53.333$ $p < 0.01$, or Condition SMFR: $Q = 50.476$ $p < 0.01$. The same pattern held for the prompt phases of these interviews: accuracy was lower in the prompt phase of Condition C than in the prompt phase of Condition SM: $Q = 20.153$ $p < 0.05$, or the Condition SMFR: $Q = 33.442$ $p < 0.01$. This result differs from that found in Experiment 7, in which no significant interaction of interview group with phase for accuracy of information elicited was discovered.

If those subjects who reported no correct information and made no errors during the readback phase are counted as having been 100% accurate, then the raw data for accuracy is as shown in Table 12.4:

	Initial	Readback	Prompt	Total
C	85.456	73.333	57.331	60.151
SM	93.083	100.000	77.484	90.189
SMFR	95.065	98.001	90.773	94.327

TABLE 12.4 - MEAN ACCURACY OF RECALL x INTERVIEW GROUP x INTERVIEW PHASE, INCLUDING DATA FOR SUBJECTS PREVIOUSLY DISCOUNTED

Total mean accuracy across interview groups and items of factual information was 92.833%. A single-sample t-test showed this to be significantly above the chance score of 33.333·%: $t(19) = 11.200$ $p < 0.01$.

Monitoring the Source of Information: These data were calculated in the same way as in Experiment 7.

	No. of subjects giving info. in the category	Mean accuracy %
Condition SM		
VO Only	10	86.667
Consistent	10	96.000
Inconsistent	4	60.000
Video Only	1	100.000
Condition SMFR		
VO Only	8	100.000
Consistent	10	100.000
Inconsistent	1	100.000
Video Only	4	100.000
Total accuracy	20	92.8·33·

TABLE 12.5 - MEAN ACCURACY SCORES FOR EACH TYPE OF FACTUAL INFORMATION x INTERVIEW GROUP, PLUS NUMBER OF SUBJECTS GIVING INFORMATION IN EACH CATEGORY

The Effect of Stereotypes on Recall: Of thirty subjects, seventeen correctly remembered that the profession of one of the characters in the videorecording was that of lawyer, and twelve correctly identified the profession of one of the characters as labourer. Of the remainder, eight misremembered another white-collar job (e.g., "accountant"), or explicitly said that the person had a white-collar job, in place of "lawyer", and five remembered no details about that profession. In the case of the labourer, eight misremembered another blue-collar job (e.g., "plumber"), or explicitly stated that the person had a blue-collar job. One subject stated that this character was involved "in

the arts". Nine subjects remembered no details about that profession.

Of those who identified one of the men as a lawyer, seven (of a possible seventeen) mistakenly identified him and the man in the suit. The remaining ten correctly identified him as the casually-dressed man. This is revealing when it is borne in mind that, in Experiment 7, 100% of those subject who correctly remembered that the profession of one of the men on the videorecording was a lawyer thought that the man in the suit was the lawyer. Similarly, of those who identified one of the men as a labourer, only five misidentified him as the casually-dressed man, the remainder correctly identifying him as the man in the suit. Again, in Experiment 7, 100% of those subjects who remembered that one of the characters was a labourer mistakenly thought that he was the casually-dressed man.

This effect is particularly interesting when broken down by interview group. This data is shown in Table 12.6, overleaf.

In each case, the proportion of correct identifications of the professions of the characters in the videotape increased from the Condition C to Condition SM, and from Condition SM to Condition SMFR. There was however no corresponding increase in the number of subjects correctly recalling the professions of the two characters.

	No. of Ss correctly remembering lawyer	Percentage identifying casually-dressed man as lawyer
Condition		
C	5	40.000
SM	7	71.429
SMFR	5	80.000
Total	17	64.706

	No. of Ss correctly remembering labourer	Percentage identifying smartly-dressed man as labourer
Condition		
C	2	50.000
SM	6	66.667
SMFR	4	100.000
Total	12	75.000

TABLE 12.6 - NUMBER OF SUBJECTS REMEMBERING AND PERCENTAGE CORRECTLY IDENTIFYING LAWYER/LABOURER ACROSS INTERVIEW CONDITIONS

Chi-square could not be performed on these data, since the expected values in three of the six cells fell below 5. Hence Fisher's Exact Probability Test was carried out. Since this test can only be performed on a 2x2 contingency table, it was necessary to combine scores obtained in Conditions SM and SMFR and compare the effect of source-monitoring instructions *per se* with that of control conditions. As far as identifying the lawyer was concerned, a result of $p = 0.1384$ was achieved. For identification of the labourer, $p = 0.056$. These results indicate that subjects who received source-monitoring instructions were not significantly better than control subjects at identifying the professions of the characters in the video-recording, even though the

trends shown in Table 12.6 lie strongly in that direction.

Discussion

With regard to the hypotheses set out in the Introduction, the following conclusions can be drawn:

1. The first hypothesis held that, in the initial phase, subjects in Condition SMFR should recall more information, make fewer errors, and be more accurate, than those in either of the other conditions.

This hypothesis does not appear to hold: any gains made by subjects in Condition SMFR over those in Conditions C and SM cannot therefore be attributed to improved performance *solely* in the initial phase, i.e., the phase during which source-monitoring instructions were presented.

2. The second hypothesis held that, in the readback phase, subjects in Condition SM should remember more information, make fewer errors, and be more accurate, than those in Condition C.

Again, this hypothesis did not hold as regards the number of errors made. Nevertheless, greater accuracy was shown, as predicted, in the prompt phase, by subjects in Condition SM than by subjects in Condition C (the prompt phase being the phase during which subjects in Condition SM received source-monitoring instructions). However, a significant difference was also discovered in the prompt phase between subjects in Condition *SMFR* and subjects in Condition C, with no significant difference between Conditions SM and SMFR.

Given that no significant difference in accuracy was discovered between subjects in Conditions C and SMFR for the *initial* phase, (see above,) it should be concluded that although source-monitoring instructions do appear to be efficacious in improving recall, this improvement cannot be ascribed to any single phase of the interview.

3. The third hypothesis held that subjects in Conditions SM and SMFR should make fewer errors for factual information than those in either of the other two conditions, and make fewer stereotype-consistent errors regarding the true professions of the characters in the video-recording.

Although no significant interaction was discovered between interview group and information type for number of errors made or accuracy of information reported, there was a significant interaction between interview group and information type for correct information reported. This applied only to visual and conversational, and not factual (as predicted) information. Neither do subjects in source-monitoring conditions appear superior to controls in identifying the correct professions of the men featured on the video-recording.

Source-monitoring interviews appear to have been particularly efficacious in reducing the impact of stereotypic expectancies regarding the professions of the two men in the video-recording. Accuracy of memory for this aspect of the to-be-remembered information increased with interview type, being

lowest in the Condition C and highest in Condition SMFR. This was probably the most important hypothesis of the present experiment. It has been found that, when presented after a delay, source-monitoring instructions are efficacious in reducing the impact of stereotypic expectancies on recall. This effect holds in the case of the basic source-monitoring interview, in which source-monitoring instructions are given during the readback phase, and even more powerfully for the source-monitoring-plus-free-recall interview, in which source-monitoring instructions are included at the beginning of the interview, in the free-recall phase.

4. Hypothesis 4 held that, overall, performance in Conditions SM and SMFR should be characterised by greater recall of correct information, greater accuracy, and fewer errors than performace in Condition C; and that recall performance in Condition SMFR should be similarly superior to that in Condition SM.

There was no significant effect of interview group on the amount of correct information recalled. The number of errors of recall made however decreased from Condition C to SMFR. More errors were made by subjects in Condition C than subjects in Condition SM, and more errors were made by subjects in Condition SM than subjects in Condition SMFR. Therefore the present experiment provides some evidence that the source-monitoring technique (particularly when

combined with free-recall) is a powerful tool for reducing the number of errors made in recall of an event. Also, the source-monitoring technique was found to increase the *accuracy* of subjects' recall of the witnessed event. The inclusion of source-monitoring instructions during the free-recall phase of the interview did not increase accuracy of recall above the level achieved by the basic source-monitoring interview however; although there was a trend in this direction.

Overall, these results provide support for the use of the source-monitoring interview in situations in which there is reason to suspect that a witness' recollections may be coloured by stereotypic information or expectancies. The source-monitoring technique should not be expected to increase the amount of correct information recalled by a witness - it should not be used as a retrieval mnemonic - but can be expected to reduce the number of errors made and consequently increase the witness' over all accuracy. These effects may hold simply through the process of forcing the witness to think more deeply about the events in question, and hence reducing their likelihood of reporting incorrect information, or information of which they are less sure. The source-monitoring technique also appears to be a powerful technique for the reduction of the effects of stereotypic information - at least, when presented after a delay. This was the area in which the beneficial effects of the interview were expected to

be most strongly apparent. Given not only the positive results of the present experiment, but also the long history of experimental success with source-monitoring tasks presented by Marcia Johnson and her colleagues, one need have little hesitation in recommending its use in situations in which social memory is likely to have been distorted by stereotypic information: and not just eyewitness memory situations.

Four differences are apparent in the pattern of results achieved in the present and the previous experiment. 1) In Experiment 7, an interaction between interview condition and phase of interview was found for correct information, but no such effect was discovered in the present experiment. This may be explicable in terms of a model in which stereotypic and other forms of information are stored together in memory immediately after exposure to stereotypic information, but the effect of stereotypic information decays across time. Hence when there is no delay between receipt of information and interviewing, a special technique such as source-monitoring may be required for access to certain pieces of information; hence these pieces of information would only be recalled in the phases during which source-monitoring was carried out. Once a delay has occurred, however, this pattern would no longer hold. 2) There was a significant effect of interview group on number of errors made in the present experiment, but not in

Experiment 7. In the present experiment, the main difference that was discovered between source-monitoring and control groups was in terms of the number of errors made. Since source-monitoring appeared efficacious when presented after a delay, but not when no delay occurred, this is the pattern of results that would be predicted. 3) Interview group had a significant effect on accuracy in the present but not in Experiment 7. This effect again was due to the reduction in number of errors made when source-monitoring instructions were given. 4) There was a significant interaction of interview group with phase of interview for accuracy in the present experiment but not in Experiment 7. This effect highlights the locus of the effect discussed under 3), above: the effect on accuracy of source-monitoring instructions seems to have been due to an effect on the number of errors made; this effect in turn appears to have been more powerful during the phases in which source-monitoring instructions were actually given.

A second aim of the present experiment was to test Wright's (1992) model of the misinformation effect: another important area of eyewitness research. Wright's claim is that, in the misinformation paradigm, of which the present experiment is a variation, while memory of the two original sources is still present in the mind, then either piece of information can be accessed independently. However, with the passage of time, memory for sources will

fade, and original information will blend with misinformation in the memory. Neither piece of information can now be accessed independently. Since memory for the source of information should have been lost, then source-monitoring should be ineffective.

The present experiment does not support this claim. Source-monitoring appears to have had very powerful effects on memory, reducing the number of errors made and increasing the accuracy of recall. Hence, the results presented here count against Wright's model. They give no cause to believe that, over time, a blend occurs between original and subsequently-presented misinformation.

A supporter of the Wright model could claim that this argument approaches circularity: that the use of a source-monitoring task presupposes that the sources of information are still discriminable in memory, and hence the success of the source-monitoring task in such an experiment proves little beyond the fact that the two information sources *have not yet* combined. However, such an argument is itself circular: the holder of such a position would be asserting nothing beyond the self-evident statement that "when sources of memory have combined in the memory, then those sources will have combined". Any experiment that indicates Wright's position to be incorrect will in fact show nothing beyond the fact that the memory sources used in that experiment have not yet combined. Such an argument would indeed be circular, and generate no experimentally-testable hypotheses.

Hence, the most optimistic assessment that can be made of the Wright model on the basis of the present experiment is that it is as yet unproven and certainly needs further work; the most pessimistic that it is, if not necessarily false, at least unhealthily close to circularity.

CHAPTER 13 - DISCUSSION

Introduction

The aim of the present thesis was to investigate the effects of stereotypic information on a number of aspects of the eyewitness situation, including judgment of and memory for a witnessed crime, and the efficacy of a source-monitoring interview in reducing the negative impact of stereotypic information. The present chapter reviews the major findings for each of the three areas and attempts to combine them to form a preliminary model of stereotypes and eyewitness memory. Implications for this theory are given in papers by Srull (1984) and Toglia, Shlechter and Chevalier (1992). These implications are also considered and incorporated into the model. The chapter concludes with a discussion of the type of work which should be conducted in future if the model is to be tested and refined further.

Judgment

Chapter 2 included a discussion of the conceptualisations proposed in the literature about the effect of stereotypic information on judgment. The most important papers in this respect are those by Bodenhausen (1988) and Macrae (in press). Both papers are strongly concerned with juridic judgments, which, as Hamilton (1979) points out, may be a special case. Few other

real-world tasks include both "on-line" and "memory-based" components (Hastie and Park, 1986). Nevertheless, it was possible to draw from the literature some predictions for the eyewitness situation. Bodenhausen (1988) identifies three possible hypotheses to account for the effects of stereotypes on judgment, and concludes that the biassed interpretation hypothesis gives the best account of his findings. According to this hypothesis, when a stereotype is presented *after* receipt of target material, no judgmental bias should be found. Only one of the three hypotheses allows for such an effect; this being the heuristic hypothesis. As Macrae (in press) remarks, there are two possible versions of this. According to one version, activation of a social stereotype provides the perceiver with a range of possible information upon which (exclusively) judgments may depend. According to the other hypothesis, pre-computed judgments may be stored in memory, and the presentation of stereotypic information will simply cause these judgments to be retrieved. Macrae also points out that Bodenhausen's means of discriminating between the biassed interpretation and selective processing hypotheses is based upon the shaky assumption of positive correlation between recall for and judgment of target information (Hastie and Park, 1986). Hence it can be argued that, although Bodenhausen's theory of the primary importance of

the biassed interpretation hypothesis may well be valid for the juridic judgment situation, it may not be generalisable to other, more common, real-world situations. It was therefore necessary in Experiments 1 - 3 to test the effects of stereotypes presented both at encoding and retrieval of target material.

Bodenhausen and Wyer (1985) provide three hypotheses to account for the circumstances under which stereotypic information will affect judgments. The impact of stereotypic information is presumed to vary with the amount of other (non-stereotypic) information provided in a given case. The heuristic hypothesis holds that non-stereotypic information will be important in making judgments only when stereotypic information is unavailable. The work of Locksley and her colleagues (Locksley, Borgida, Brekke and Hepburn, 1980; Locksley, Hepburn and Ortiz, 1982; Locksley, Ortiz and Hepburn, 1980), while controversial in other respects, (Grant and Holmes, 1981) appears to refute this hypothesis. At the opposite extreme, the default hypothesis holds that stereotypes are used only as a last resort in judgment tasks. While Locksley's work appears to confirm this hypothesis, closer consideration implies that it in fact appears to support the integration hypothesis, as proposed by Bodenhausen and Lichtenstein (1987). According to this hypothesis, subjects consider both stereotypic and

non-stereotypic information in reaching a judgment. The relative weight attached to each may be to some extent situation-specific, according to the complexity of task demands. It was argued that this conclusion could be integrated with that proposed by Devine (1989a and b), who argued that even a non-bigot might occasionally make a judgment based upon stereotypes when task demands are particularly high. Unfortunately, such a model, although highly intuitively appealing and reconcilable with the results reported in the literature, cannot be used to generate testable hypotheses; hence the results achieved in Experiments 1 - 3, although quite compatible with the model, cannot really be considered direct evidence in its favour.

While Experiments 2 and 3 involved both memory and judgment components, the main thrust of this section of the thesis was to test a) the impact of stereotypic information in paradigms with direct implications for the eyewitness memory situation, and b) to what degree this impact might vary with time of presentation of stereotypic information. This manipulation would provide a test of the "heuristic hypothesis" proposed by Bodenhausen (1988), and also, indirectly, the relevance of work of juridic judgments to the present situation. Naturally, this work could not be held to have any powerful implications for the hypotheses proposed by Bodenhausen or Macrae;

however a powerful effect of stereotypes on judgment for a complex situation like the eyewitness situation would also be predicted by the "model" proposed by Srull and Wyer (1989) for the less specialised area of impression-formation. A demonstration of such effects would therefore tend to support the Srull and Wyer position.

Experiment 1 demonstrated an effect of stereotypic information on judgment of an ambiguous action. When a protagonist was perceived to be larger than his victim, an ambiguous shove administered by the protagonist was judged in more negative terms than when he was smaller than or of an equal size to his victim. This demonstrated that stereotypic does indeed appear to have a powerful effect in a situation which is more closely analogous to eyewitnessing situation than is the kind of juridic task employed in much earlier research. This effect was demonstrated even where the absolute difference in body-size of the two actors was quite small: an actual difference of about three inches and twenty-five pounds. When such effects could be achieved using actors whose body-sizes differed so slightly, it was argued that stronger effects would occur where body-sizes differed more dramatically; as could well be the case in the real world.

While Experiments 2 and 3 contained elements of both memory and judgment tasks, neither one showed

a differential effect with stereotype presentation time; indeed, in both cases, data from the two presentation-time conditions was combined for the purpose of analysis. Bodenhausen's (1988) study demonstrates that, in the juridic task, a stereotype presented *after* target material is unlikely to affect judgments based upon that material. The implication of these studies however is that, in a more realistic task, in which judgments tend to be more on-line than memory based, stereotypic information does affect judgments.

In experiments even as marginally realistic as Experiments 2 and 3 above, the heuristic hypothesis does appear to have explanatory power. Indeed, neither the biassed interpretation hypothesis nor the selective processing hypothesis is equipped to deal with such results. This strengthens the case for claiming that the juridic task represents a special situation in the stereotypes and judgment literature, and that other real-world situations deserve attention in their own right.

Of course, it is impossible to discriminate on the basis of these three sets of results between the two versions of the heuristic hypothesis proposed by Macrae (in press). It is impossible to tell whether the presentation of a stereotype activated stereotypically-linked information *upon which* subsequent judgments were based, or whether

it activated pre-computed judgments. This is a question for future research to address: however the suggestion could be tentatively advanced that the answer is likely to be situation-specific. Each means of producing judgments has an associated advantage and disadvantage; each involves a trade-off between accuracy and ease of judgment. Where more cognitive resources are available, one might predict stereotypic information to be generated and judgments to be made on the basis of that; where fewer resources are available, ease will be at a premium and a pre-computed judgment is likely to be produced.

In summary of the conclusions gained concerning stereotypes and judgment then, it can be said that a) the integration hypothesis appears likely to give the best indication of *when* judgments will be made on the basis of stereotypic information. The relative use of stereotypic and "other" information is likely to be very situation-specific; where there are greater demands upon cognitive resources, then reliance upon stereotypic information appears likely to be greater b) the heuristic hypothesis is the only one of the three outlined by Bodenhausen (1988) which can account for all of the results reported here. Although Bodenhausen's own results run counter to this hypothesis, this is explicable by the fact that Bodenhausen employed *juridic* tasks which appear likely to be special cases as far as

real-world tasks are concerned. The effect of stereotypes presented at retrieval also has great ramifications for the stereotypes-and-memory field, addressed in the next sub-section, below.

Finally it should be mentioned that this area is unlikely to develop much beyond its present state until more adequate theoretical foundations are built. Science progresses through the testing and refinement of hypotheses. The hypotheses in this area are however extremely limited: applying as they do for the most part to the juridic situation. The only theoretical formulation of the effects of stereotypes on judgment outside this narrow area of focus is that by Srull and Wyer (1989); and this paper addresses only the slightly less narrow area of impression-formation. Some theoretical formulation of the effects of stereotypes on judgment in other situations, and using other methodologies - particularly the crime situation - is really necessary for the advancement of work in this area.

Memory and Interview Techniques

Memory: The present sub-section consists of a summary of the results of the experiments reported in the present thesis which tested the effects of stereotypes on memory, and a discussion of the inferences which can be drawn from these. In a separate sub-section, below, a preliminary model

of the effects of stereotypes in the eyewitness memory situation is presented.

The effect of stereotypic information on memory is expected to differ with both time of stereotype presentation and type of material (Bodenhausen and Lichtenstein, 1987; Srull, 1984; Snyder, 1984; Wyer, Bodenhausen and Srull, 1984). The present thesis considered the effects of stereotypes presented at encoding and at retrieval, and also different types of target material. Both variables must be considered in attempting to give a coherent account of the results achieved.

Experiments 2 and 3, above, employed stimuli that are of use in person identification. Stereotypic information was presented both at encoding and at retrieval. In neither case was an effect of time of stereotype presentation discovered. Of four hypotheses generated in Experiment 2, however, only one was fulfilled: when subjects were told that a target committed the stereotypically aggressive crime of assault, a larger body-shape was regenerated than when they were told that the target committed a crime that was not stereotypically linked with aggressiveness. Hence it may not be valid to draw too many conclusions from this study. Experiment 3 looked at the regeneration of face stimuli. It was found that when subjects were told a target person received a 2i degree, subsequent regenerations of the target face were later judged

"more attractive" than when subjects were told that the target received a 2ii degree. When subjects were told that the target committed an assault, regenerated faces were judged more aggressive and dishonest looking than when subjects were told that the target committed a theft. A comparison of the results of Experiments 2 and 3 implies that faces may be more susceptible to the action of social stereotypes than are bodies. This issue is discussed in more detail below.

Later experiments examining the effects of stereotypic information presented at retrieval failed to find any powerful effects, however. In Experiments 4 and 5, biographical target information was used to test the rival claims of Snyder and Uranowitz (1978c) and Bellezza and Bower (1981) about the effects of stereotypic information presented at retrieval. These studies produced mixed evidence, but on the whole it appears unlikely that stereotypic information will usually act as retrieval schemata for target material. When recognition testing took place immediately after receipt of target material, no difference was discovered in the number of stereotype-consistent "hits" generated across conditions. When testing took place after a delay, indeed (Experiment 5), more stereotype-inconsistent than stereotype-consistent information was generated in the stereotype

condition. There was no difference between conditions in amount of stereotype-inconsistent material generated, however. Hence this cannot be written off as a "simple" inconsistency effect like that discovered by Hastie and Kumar (1979), and held by the "Srull-Hastie model" (Hastie, 1980; Srull, 1981) to be the most basic effect in the literature. Only in the recall paradigm of Experiment 4 was the hypothesized consistency effect actually discovered; reinforcing Srull's (1984) contention that recall and recognition paradigms may sometimes produce different results.

Recall is generally the more relevant paradigm in the present area. While recognition memory may be used by the eyewitness in, e.g., identity parades, or when looking through collections of mug-shots, recall memory will be of far greater use in the police interview. This was tested in Experiment 6, in which stereotypic information was presented at retrieval of a dramatic staged incident. The material used in this experiment perhaps had more direct relevance to the eyewitness situation than body/face or biographical information. No effect of stereotypic information was found here.

Taking these experiments together, the effect of stereotypic information presented at retrieval does appear to vary with target material used. Although there is little reason to hold that they act as retrieval schemata - for any type of

material - Experiments 2 and 3 appear to show that stereotypes presented at retrieval can have effects very similar to those of stereotypes presented at encoding. In Experiment 6 however, few significant effects were achieved when stereotypic information was presented at the retrieval of material from a dramatic staged incident. It could be argued that the experiments reported in the present thesis vary in the level of analysis which is employed. The relevance of this factor can be seen when the meta-analysis of the literature, presented in Chapter 6, above, is taken into consideration. Most of the studies reviewed consider social memory in the sense of person- or group- trait or behaviour information. While this is applicable to Experiments 4 - 6, it is less applicable to Experiments 2 and 3.

It is possible to apply Stangor and Ruble's (1989) model to this argument: Conceivably, where person- or group- trait or behaviour information is used for target materials in an experiment, then subjects' expectations will be more powerful than where materials such as face or body stimuli are used. By this reasoning, a consistency effect of stereotypic information is most likely in Experiments 4 - 6 in the present thesis, whereas in Experiments 2 and 3 different effects might be expected. Time of stereotype presentation in these experiments may therefore be of less

importance than the characteristics of the materials considered.

In summary, then, the effects of stereotypic information presented at the retrieval of person- or group- trait or behaviour information may not be very powerful in the genuine eyewitness situation. This positive comment may be offset by a negative one however: In such a situation, it is unlikely that stereotypic information could be used as retrieval schemata in the way that Snyder and Uranowitz (1978c) suggest. The role of stereotypic information presented at retrieval of information which lies at a different level of analysis - in this case, face and body information, which may prove of especial relevance to the police - may be more important. While there appears once again to be no reason to predict a retrieval-schema effect (faces and bodies regenerated in the two conditions in these experiments did not differ in *likeness* to the target stimuli) neither is there any reason to predict lower effects when stereotypes are presented at retrieval than when they are presented at encoding. Hence, biasses in regeneration may occur when stereotypic information is presented at retrieval: e.g., during discussions of the incident between the witness and his/her friends, or during the police interview itself.

To extend this line of reasoning, models such as the "Srull-Hastie model" and its later incarnations may indeed not be wholly applicable to real-world situations, since they appear to be based to a great extent on experiments which employ material quite unlike what is usually encountered outside the laboratory (cf. Ruble and Stangor, 1986). Hence the "Srull-Hastie model" - in both its original and later versions - may prove inapplicable to at least the majority of real-world situations. It is however conceivable that an inconsistency effect might be discovered in situations where expectancies are moderately low. Such a formulation would however encounter problems in predicting under what particular conditions expectancies might be high or low. There would also be the theoretical problem of specifying exactly what kind of social stereotype it might be that would generate *low* expectancies.

Support for the Stangor and Ruble model therefore appears quite convincing. The model is able to explain not only the divergence of results in the social memory literature as a whole, but also those specifically discovered in the present thesis. From this model one would draw the prediction that, in the "average" eyewitnessing situation, where the witness enters the situation with cognitive expectancies such that s/he encodes material in terms of person- or group- trait or behaviour stereotypes, then, on subsequent recall,

a *consistency* effect is most likely. This will probably manifest itself in terms of superior recall for stereotype-consistent material; however, it is also possible that it will engender a degree of response-bias (see Experiment 4, above,) causing a larger number of stereotype-consistent errors ("false alarms") to be made. Depending on the point in the police investigation at which the interview takes place, this could be more or less important. Where the police are trying to generate leads in a difficult criminal investigation, the generation of a certain number of "false alarms" may not be of great consequence (Geiselman, Fisher *et al*, 1985); however, where harder evidence is required the implications may be more negative.

Interview Strategies: It was argued in Chapter 10 that although a great deal of work in the psychological literature has addressed the issue of the enhancement of eyewitness memory through the use of different interview strategies - especially hypnosis and the cognitive interview - none of these strategies is wholly appropriate to the task of reducing the negative impact of stereotypic information on memory. Such an effect may be particularly likely where the witnessed event is encoded in terms of pre-existing social stereotypes (see the previous sub-section), and the stimulus material is of the type that is most

likely to be encountered during the eyewitnessing situation - i.e., person- or group- trait or behaviour information. One might do well to predict distortions in eyewitness memory to form a *consistency* effect. A crime encoded in terms of a particular social stereotype appears likely to be remembered as having more closely matched the contents of that stereotype than in fact it did. The technique of source-monitoring developed by Marcia Johnson and her colleagues (e.g., Johnson, 1977; Johnson, Raye and Durso, 1980; Johnson, Taylor and Raye, 1977; Johnson, Raye, Foley and Foley, 1981; Lindsay and Johnson, 1989,) seemed a promising means of discriminating between memories whose source lies in the witnessed incident itself and those whose source lies in the social stereotype in terms of which the incident was encoded.

Two experiments were reported in which the efficacy of the source-monitoring interview was examined. Both involved interviewing subjects about a previously-viewed video-recording which involved an encounter between two men in a bar, and discrimination of the actual professions of these two men.

In the first experiment, the source-monitoring interview, presented without a delay between encoding and retrieval, was not found to increase the amount or accuracy of information recalled, decrease the number of errors made, or reduce the

impact of stereotypic information on memory relative to a control interview. With the addition of the externally valid measure of a delay between encoding and retrieval, however, the source-monitoring interview proved more effective.

This certainly seems a promising means for the reduction of stereotypic biasses in eyewitness memory. Future work would do well to continue the investigation of the source-monitoring technique, employing not only the methodological approach used in Chapters 11 and 12, but also a methodology which might offer less experimental control, but more clear genealisability to the real-world eyewitnessing situation, e.g., a video-recording of a crime in progress, or a staged incident.

The practical implications of this technique may prove limited, however. Since it does not claim to be a useful technique when the witness' memory has *not* been distorted by stereotypic information, use of the source-monitoring interview will be advantageous in only a limited number of real-world cases. This, coupled with the obvious practical difficulties involved with knowing exactly *when* a witness' memory is likely to have been subject to stereotypic bias, mean that the adoption of the technique on a large scale appears unlikely, however successful it may prove in reducing the negative impact of stereotypic information. The police service itself has traditionally taken a rather negative attitude

towards interview training; considering experience to be the best (and indeed only) teacher (see, e.g., Goodsall, 1974). This is remarkable given the importance attached by police officers to eyewitness evidence (Roy, 1991, p 5; Orne *et al*, 1984, p 171; Geiselman *et al*, 1984), and the success with which interview training has met in employment and personnel fields (see, e.g., Schuh, 1981). The outlook may be growing less bleak however. With the recent rise in interest among psychologists in police practice (see, e.e., Irving and Hilgendorf, 1980; Farrington, 1981; Softley, Brown, Forde, Mair and Moxon, 1980), there has been a concomitant rise in certain types of police training. Spivey (unpublished) discusses - and is quite positive about - the effectiveness of "social skills training" in the police forces of Britain. Adams (1985) discusses the rationale, method and procedure that lie behind "communication skills" training amongst new recruits to the police forces of the East Midlands. The success of such measures may encourage more widespread implementation of interview training throughout Britain. "...[E]veryone now wants training" asserts E Shepherd (1991, p 7). While this may be an exaggeration, the future of police interview training certainly appears brighter now than previously. The main obstacle, as E Shepherd

pointed out in an earlier paper, may be resistance within the profession itself (E Shepherd, 1985).

The source-monitoring technique *has* proven effective in combatting the misinformation effect (Lindsay and Johnson 1989; 1991 a, b, & c). It is perhaps to be hoped that it might at some time in the future be adopted by law-enforcement bodies with a stronger interest in interviewing techniques than they have at present, on the strength of this fact, and that its use may then spill over into cases in which witness' memories are believed to have been biassed by stereotypic information.

Towards a Model of the Effects of Stereotypes in the Eyewitness Situation

Throughout the thesis, and above, a preliminary conceptualisation of the effects of stereotypes upon social (and specifically eyewitness) memory has been developing. This sub-section attempts to bring together the arguments that have so far been presented and to summarise them in the form of a preliminary model of the effects of stereotypes in the eyewitness situation.

The question has many facets. Hard-and-fast conclusions regarding this complex area will almost certainly not be reached for some years, when every aspect of the question has received due consideration. The present thesis has considered effects of stereotypes presented both at encoding

and at retrieval, with or without delay conditions, and for different types of target material. From this some preliminary remarks may be made regarding stereotypic effects in the eyewitnessing situation: particularly given the results of Experiments 7 and 8. The aim here is not so much to try to construct a final model of the effects of stereotypic information on eyewitness memory, as to give an outline of which factors may be important, and when.

The Effect of Time of Stereotype Presentation:
Stereotypic information can affect memory whether presented at encoding or at retrieval (Experiments 2 - 5, above). However, as argued above, the effects discovered may vary with the level of analysis employed. Stereotypes presented at retrieval appear likely to have an effect when information is analysed at the level of e.g., face or body information, and there is no delay before testing takes place. Experiments 2 and 3 showed that the effects of stereotypes presented at retrieval were not statistically different from effects of stereotypes presented at encoding when both of these conditions held. Experiment 4 used a different level of analysis, employing for target materials a biography of a young woman. Again, effects of stereotypic information presented at retrieval were discovered. While these effects appear to have been more pronounced

when a free-recall paradigm rather than a recognition paradigm was employed (cf. Srull, 1984), neither paradigm produced very powerful results. The range of experiments reported in the present thesis does not allow many inferences to be drawn regarding the effects of testing procedure, however. This is an area which future research might do well to address. In Experiment 5, the same materials and level of analysis were employed, but when a delay condition was included, the results achieved contradicted those of Experiment 4. Experiment 6 also employed a delay condition. No effects of stereotypic information presented at retrieval were discovered in this experiment.

Hence, as argued above, in the real world, the effect of stereotypes presented *after* target information (whether during discussions with friends or during the police interview,) may not be very large. The majority of information contained in the crime situation is likely to be categorisable as person- or group- trait or behaviour information; also, some degree of delay is likely between witnessing a crime and discussing it with friends or a police officer. The exact length of time which will prove significant cannot be specified here since this was not an aim of any of the experiments reported; however one could suggest that the exact length of time which is important may differ with individual

cases, but is unlikely to exceed three weeks (Experiment 5).

The basic paradigm of Experiments 7 and 8 was to present stereotypic information at encoding. Every experiment in the present thesis in which stereotypes were presented at encoding showed an effect. Effects are perhaps likely to be similar to those of stereotypes presented at retrieval in that they will vary with the level of analysis employed, being less pronounced when information is analysed at the level of the facial or body characteristics of a target, and more pronounced when analysed at the level of person- or group-trait or behaviour information. Experiments 7 and 8 demonstrate that effects are also likely to be more pronounced when testing is immediate than when it is delayed. The efficacy of the source-monitoring interview in reducing the impact of stereotypes *only* after a delay implies (as suggested in Chapter 11) that the two sources of information become discriminable only after a period of time has elapsed (*pace* Wright, 1992). The impact of stereotypic information therefore appears likely to decay with time.

The Effects of Different Testing Paradigms: The testing paradigm used appears likely to affect the quality of information generated in an eyewitness memory experiment. The present thesis, being primarily concerned with outlining possible

effects in a little-explored area of research, did not explicitly consider the effects of different paradigms; hence the preliminary model given above makes little mention of this variable. However it is certain that any more comprehensive model which may later be developed must take this variable into account. The present sub-section gives a brief account of conclusions reached in the experimental situation by Toglia, Shlechter and Chevalier (1992) and in a literature review by Srull (1984), to outline the kinds of effects that appear most likely to occur.

Research into the effects of different modalities on eyewitness memory has been necessary for number of years. Yuille and Cutshall (1986) criticise much earlier research for attempting to generalise from ill-controlled and often unrealistic laboratory experiments to real-world situations in which memory may be better in all respects. Indeed, inter-paradigm differences may be more subtle even than this: memory for different types of material may vary across paradigms in different ways (Yuille and Cutshall, 1984).

As Toglia *et al* point out, there are but a few studies of modality effects relevant to eyewitness research, and these have failed to show a consistent pattern of effects. Brown, Heyman, Preshill, Rubin and Wuletich (1977) found no memory differences between subjects who had

witnessed a live event and those who had simply heard a description of the event. Wagenaar and Visser (1979), however, found superior recall for spoken-only than for television (sound-plus-vision) weather forecasts. Lassen (1988) found that indirectly experienced events appeared more difficult to remember than directly experienced events: a result which appears to contradict that of Wagenaar and Visser.

However the most systematic study of these effects is certainly that by Toglia *et al* themselves, and their conclusions stand repeating in full.

They report two studies, in which a staged event was experienced either directly, by the subjects themselves, or indirectly, through verbal reporting. Immediately following or two weeks after experiencing the incident, subjects' memories were tested for different types of information from the incident. Memory decayed less with time in both the free-recall and recognition paradigm (cf. Experiment 6, above) for those who had directly experienced the event than for those who had indirectly experienced the event. It was also found that *action information* was better retained than conversational or descriptive information, and also that retention for different types of information varied with exposure type. In both conditions, action details were remembered relatively well, but recall levels

were higher for the indirectly experienced event as conversation and physical description information were better retained. These results were found in the absence of any systematic between-condition difference in terms of the rated authenticity and stressfulness of the event, or confidence in the responses.

Of particular importance to the present thesis was the study of the effects of *delay* upon memory. Although this is a variable with great external validity, it has received little attention in the eyewitness literature. In the present situation there was a second important reason to study the effects of delay: it was believed that the effects of stereotypic information on memory might change with time (Wright, 1992). This is, therefore, a particularly important variable for any future work in this area to consider. The findings of Toglia *et al* are of particular importance here in that they demonstrate that the effects of delay are likely to vary not only with stereotype condition, but also with testing paradigm employed and type of information generated. Ideally, future research in the present area should take account of all of these variables in an attempt to develop a model which specifies effects under all possible conditions. Failing this, experiments should at least be designed with these effects in mind, and extrapolations from data thus generated should be limited to paradigms under which they

might reasonably be expected to hold. Toglia *et al* appear to find relatively little difference between recall and recognition testing paradigms. Hence, extrapolating from this result, it would appear that future research need not worry overmuch about which particular testing paradigm to use, but to use whichever is more convenient.

Srull (1984) however considers this question in more detail. He points out that the social cognition literature includes very few explicit comparisons of recall and recognition paradigms, but that nevertheless important theoretical differences exist between the paradigms, and these should be borne in mind when such experiments are designed.

The use of the free-recall testing procedure has one major drawback as far as research in the present area is concerned: it is one of the least sensitive measures of memory available. Srull holds therefore that it is a poor method for the investigation of small but theoretically important differences. This is most problematic when long delay intervals are used. On these grounds therefore, the free-recall measure appears inappropriate to the present situation: as argued above, only subtle effects of stereotypic information on eyewitness memory are to be expected; and the use of delay intervals is an important aspect of the work. Nevertheless, it should be borne in mind that the free-recall

paradigm is the most appropriate to present circumstances (since the majority of information that eyewitnesses produce comes during the police interview, during which a free-recall measure is most likely to be used) and also that Experiments 4, 6, 7, and 8 above *did* isolate stereotypic effects on memory even when free-recall measures were used.

The recognition paradigm is, Srull points out, more sensitive than the free-recall paradigm to the retention of virtually any type of information. In most instances, the majority of drawbacks that can be associated with use of this paradigm can be eliminated by careful construction of the test itself (e.g., making sure there is no *a priori* basis for discriminating between familiar and novel items; controlling the similarity of distractors and test items, etc.) However the main disadvantage of the recognition paradigm in the present situation is its lack of ecological validity; a drawback which cannot really be circumvented.

Bearing these factors in mind, it is difficult to suggest which paradigm should be used in future research. Each approach has advantages which compensate to some extent for the disadvantages of the other. Perhaps the best advice is that future research should, wherever practicable, use both paradigms, in an attempt to check whether effects obtained are generalisable from one to the other.

Where this is impossible, the likely differences between the paradigms should be borne in mind, and some discussion given of how likely the results are to obtainin the opposite paradigm, and, based on this, of the real-world applicability of these results.

Directions for Future Research

The present sub-section attempts briefly to recap the areas which should be addressed in future research, and the way in which this should be approached.

The most obvious limitation of the research presented here is in the area of stereotypic effects on social judgments. While the research presented demonstrated powerful effects of stereotypes on judgments (particularly in Experiment 1), only a limited range of experimental methods and target stimuli was employed. Future work should address the issue of the effects of stereotypic information upon judgment where the experimental materials used have more powerful real-world implications e.g., a video-recording of a crime in progress, or a staged incident.

It was found here that the effects of stereotypes presented at retrieval are unlikely to be very powerful except perhaps under very specific sets of circumstances. While these circumstances could stand further investigation

and delineation, there would be greater utility in the further investigation of areas in which stereotypes are likely to have more powerful effects. Of particular importance would be the construction of a model of the effects of stereotypic information presented at encoding, and how these effects vary with condition. The compatibility of such a model with other models in the social cognition literature (and especially the "Srull-Hastie" model and the Stangor and Ruble model) should also be investigated.

The model presented by Stangor and Ruble (1989) appears to give the best account of the results achieved in the present thesis, and might, therefore, ultimately prove the best model to account for the effects of stereotypic information on eyewitness memory. This does however require future work to include rigourous testing of the role of subjects' expectancies on entering the eyewitness situation.

Delay interval is another important factor which should be considered in future work. The present thesis suggests that the impact of stereotypic information is likely to decrease with time. Future work should consider how long a delay is necessary for this to occur, and also whether such a model could be made compatible with research into the misinformation effect which holds that the sources of different items of information in

long-term memory will become *less* rather than *more* discriminable after a delay.

Finally, the source-monitoring interview (Chapters 10 - 12) seems a particularly promising avenue for future research, appearing when presented after a delay to be an effective means of reducing the impact of stereotypic information. The technique also appears to have great credibility given the volume of successful research conducted by Marcia Johnson and her colleagues over the past fifteen years. However, the experiments reported in the present thesis (as well as a preliminary experiment by Sussman, 1992) used a somewhat limited range of stimulus materials. Future research should consider target events with greater external validity, as well as more forensically relevant stereotypes, such as those employed in Experiments 2 and 3 above. As mentioned above, the limited range of usefulness of the source-monitoring interview technique may mean that it is unlikely to be adopted in a widespread fashion no matter how successful it may prove.

Future research might also address the question of exactly *when* in the police interview stereotypic distortions of memory are most likely to arise. Previous research (e.g., Hilgard and Loftus, 1979) suggests that during the free-recall component of a police interview, a witness' recollections are likely to be highly accurate, if

incomplete. This rule might also apply to stereotypic memory distortions: they may be less likely to occur during the initial, free-recall phase of the interview, but more likely to occur in response to direct questions on the part of the interviewer. If one could isolate the points at which such distortions are most likely to occur, then source-monitoring instructions might successfully be incorporated only at those points in an interview. Such an approach might prove to be the most efficient application of this interview technique.

BIBLIOGRAPHY

Adams G R (1977) Physical attractiveness research: Toward a developmental social psychology of beauty *Human Development*, 20, 217 - 239

Adams J (1985) *Communication skills training with new recruits* NATO Advanced Study Institute, Skiathos, Greece

Adorno T W, Frankel-Brunswik E, Levinson D J & Sanford R N (1950) *The Authoritarian Personality* Harper & Rowe, New York

Alba J W & Hasher L (1983) Is memory schematic? *Psychological Bulletin*, 93(2), 203 - 231

Allen V L & Wilder D A (1979) Group categorization and attribution of belief similarity *Small Group Behavior*, 10, 73 - 80

Allport G W & Postman L (1945) The basic psychology of rumor *Transactions of the New York Academy of Sciences* 8, 61 - 81

Allport G W & Postman L (1947) *The Psychology of Rumor* Holt, New York

Allport G W (1937) *Personality: A Psychological Interpretation* Constable, London

Allport G W (1954) *The Nature of Prejudice* Cambridge, Mass. Addison-Wesley

Anderson J R & Bower G H (1972) Recognition and retrieval processes in free recall *Psychological Review* 79, 97 - 123

Anderson N H & Hubert S (1963) Effects of concomitant verbal recall on order effects in personality impression formation *Journal of Verbal Learning and Verbal Behavior* 2 379 - 391

Anderson R C & Pichert J W (1978) Recall of previously unrecallable information following a shift in perspective *Journal of Verbal Learning and Verbal Behavior*, 17, 1 - 12

Argyle M & Cook M (1975) The socialization of altruistic and sympathetic behaviour: some theroetical and experimental analyses In *Altruism and Helping Behavior*, J Macauley & L Berkowitz (eds.) Academic Press, New York

Argyle M (1967) *The Psychology of Interpersonal Behaviour* Cox & Wyman, London

Arkes H R & Harkness A P (1980) Effect of making a diagnosis on subsequent recognition of symptoms *Journal of Experimental Psychology: Human Learning and Memory*, 6(5), 568 - 575

Aschenbrenner K M & Schaeffer R E (1980) Minimal group situations: Comments on a mathematical model and on the research paradigm *European Journal of Social Psychology*, 10, 389 - 398

Ashmore R D & Del Boca F K (1981) Conceptual approaches to stereotypes and stereotyping In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Baddeley A (1990) *Human Memory* Lawrence Erlbaum Associates: London

Bales R F (1970) *Personality and Interpersonal Behavior* Holt Rinehart Winston: New York

Baxter J S (1990) The suggestibility of child witnesses: A review *Applied Cognitive Psychology*, 4(5), 393 - 407

Bekerian D A and Bowers J M (1983) Eyewitness testimony: Were we misled? *Journal of Experimental Psychology: Learning, Memory and Cognition*, 9(1), 139 - 145

Bellezza F S and Bower G H (1981) Person stereotypes and memory for people *Journal of Personality and Social Psychology*, 41(5), 856 - 865

Belli R F, Windschitl P D, McCarthy T T & Winfrey S E (1992) Detecting memory impairment with a modified test procedure: Manipulating retention interval with centrally presented event items *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 356 - 367

Berry D S & MacArthur L Z (1986) Perceiving character in faces: the impact of age-related craniofacial changes on social perception *Psychological Bulletin*, 100, 3 - 18

Berscheid E & Walster E (1974) Physical attractiveness In L Berkowitz (ed.) *Advances in Experimental Social Psychology* Vol 7 New York: Academic Press

Billig M & Tajfel H (1973) Social categorization and similarity in intergroup behaviour *European Journal of Social Psychology*, 3, 27 - 52

Billig M (1976) *Social Psychology and Intergroup Relations* Academic press, London

Bodenhausen G V & Kramer G P (in press) Social stereotypes and memorial accuracy in criminal justice contexts In M D MacLeod & C N Macrae (eds.) *Stereotypes in the Criminal Justice System*, Falmer Press, UK

Bodenhausen G V & Lichtenstein M (1987) Social stereotypes and information processing strategies: The impact of task complexity *Journal of Personality and Social Psychology*, 52, 871 - 880

Bodenhausen G V & Wyer R S (1985) Effects of stereotypes on decision-making and information-processing strategies *Journal of Personality and Social Psychology*, 48(2), 267-282

Bodenhausen G V (1988) Stereotypic biasses in social decision-making and memory: Testing process models of stereotype use *Journal of Personality and Social Psychology*, 55(5), 726 - 737

Bodenhausen G V (1990) Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination *Psychological Science*, 1, 319 - 322

Boon J C W & Davies G M (1987a) Attitudinal influences on witness memory: Fact and fiction In Gruneberg, Morris and Sykes (eds) *Practical Aspects of Memory*

Boon J C W & Davies G M (1987b) Rumours greatly exaggerated: Allport and Postman's apocryphal study *Canadian Journal of Behavioural Science* 19(4)

Bothwell R K, Deffenbacker K A & Brigham J C (1987) Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited *Journal of Applied Psychology*, 72, 691 - 695

Bower G H, Black J B & Turner T J (1979) Scripts in memory for text *Cognitive Psychology*, 11, 177 - 220

Bower G H & Karlin M B (1974) Depth of processing pictures of faces and recognition memory *Journal of Experimental Psychology*, 103(4), 751 - 757

Brewer M B (1979) In-group bias and the minimal intergroup situation: A cognitive-motivational analysis *Psychological Bulletin*, 86, 307 - 324

Brigham J C (1971) Ethnic stereotypes *Psychological Bulletin*, 76, 15 - 38

Brooks K & Siegal M (1991) Children as eyewitnesses: Memory, suggestibility and credibility *Australian Psychologist*, 26(2), 84 - 88

Brown I, Heymann S, Preskill B, Rubin D & Wuletich T (1977) Leading questions and the eywitness report of a live and a described event *Psychological Reports*, 40, 1041 - 1042

Brown R & Kulik J (1977) Flashbulb memories *Cognition*, 5, 73 - 99

Brown R (1988) Intergroup relations In M Hewstone, W Stroebe. J Codol & G M Stephenson (eds) *Introduction to Social Psychology*, Basil Blackwell, Oxford, UK

Brown R J, Tajfel H & Turner J C (1980) Minimal group situations and intergroup discrimination: Comments on the paper by Aschenbrenner and Schaeffer *European Journal of Social Psychology*, 10, 399 - 414

Bruce V & Young A (1986) Understanding face recognition *British Journal of Psychology*, 79, 471 - 491

Bruce V (1988) *Recognising Faces* Lawrence Erlbaum Associates, Hove & London, UK

Bruce V (ed.) (1991) *Face Recognition* European Journal of Cognitive Psychology Special Issue Lawrence Erlbaum Associates, Hove & London, UK

Brunswik E (1956) *Perception and the representativeness design of psychological experiments* Berkley: University of California Press

Buckhout R (1974) Eyewitness testimony *Scientific American*, 231(6), 23 - 31

Bull R H C & Green J (1980) The relationship between physical appearance and criminality *Medicine, Science and the Law*, 20(2), 79 - 83

Burgess A W & Holstrom L L (1978) Interviewing young victims In A W Burgess (ed.) *Sexual Assault of Children and Adolescents*: Lexington

Cahill D & Mingay D J (1986) Leading questions and the police interview *Policing*, Autumn, 212 - 224

Cannell C F & Kahn R L (1954) The collection of data by interviewing In L Festinger and D Katz (eds.) *Research Methods in the Behavioural Sciences* Staples Press Ltd.: London

Cantor D & Mischel (1977) Traits as prototypes *Journal of Personality and Social Psychology*, 35(1), 38 - 48

Cash T F (1981) Physical attractiveness: An annotated bibliography of theory and research in the behavioural sciences *JSAS Catalog of Selected Documents in Psychology*, 11, Ms 2370

Cervantes M (1605/1870) *Don Quixote de la Mancha* M C Barrett, London

Chance J E & Goldstein A G (1984) Face-recognition memory: Implications for children's eyewitness testimony *Journal of Social Issues*, 40, 69 - 85

Chapman L J (1967) Illusory correlation in observational report *Journal of Verbal Learning and Verbal Behavior*, 6, 151 - 155

Christiaansen R E & Ochalek K (1983) Editing misleading information from memory: Evidence for the co-existence of original and post-event information *Memory and Cognition*, 11, 467 - 475

Clark L F & Woll S B (1981) Stereotype biasses: A reconstructive analysis of their role in reconstructive memory *Journal of Personality and Social Psychology*, 41(6), 1064 - 77

Clark N K, Stephenson G M and Kniveton B H (1986) Script-writers or story-tellers? The consequences of divergent recall strategies for police officers and students for presentation at BPS Annual Conference, Sheffield

Clifford B R (1984) Eyewitness testimony: The bridging of the credibility gap In D P Farrington, K Hawkins, and G Lloyd-Bostock (eds) *Psychology, Law and Legal Processes* MacMillan, London

Clifford B R & Bull R (1978) *The Psychology of Person Identification* Routledge and Kegan-Paul, London

Clifford B R & Richards V J (1977) Comparison of recall by policemen and civilians under conditions of long and short duration of exposure *Perceptual and Motor Skills*, 45, 503 - 512

Clifford B R & Scott J (1978) Individual and situational factors in eyewitness testimony *Journal of Applied Psychology*, 63, 352 - 359

Clifford M and Walster E (1973) The effect of physical attractiveness on teacher expectations *Sociology of Education*, 46, 248 - 258

Cohen A (1977) The cognitive basis of stereotyping: An information-processing approach to social perception *Dissertation Abstracts International* 1977 (Jul) 38(1-B) 412 - 413

Cole C B & Loftus E F (1989) The memory of children In S J Ceci, M P Toglia and D F Ross (eds) *Children's Eyewitness Testimony* Springer - Verlag, New York.

Cook M (1978) Eye movements during recognition of faces In *Practical Aspects of Memory*, eds. M M Gruneberg, P E Morris and R N Sykes Academic Press, London

Cook S W (1939) The judgement of intelligence from photographs *Journal of Abnornmal and Social Psychology*, 34, 384 - 389

Cortés J B & Gatti F M (1965) Physique and self-description of temperament *Journal of Consulting Psychology*, 29, 432 - 439

Crocker J, Hannah D B & Weber R (1983) Person memory and causal attributions *Journal of Personality and Social Psychology*, 44, 55 - 66

Cross J, Cross J & Daly J (1971) Sex, race, age and beauty as factors in recognition of faces *Perception and Psychophysics*, 10, 393 - 396

Cunningham J L (1988) Contributions to the history of psychology: XLVI. The pioneer work of Alfred Binet on children as eyewitnesses *Psychological Reports*, 62, 271 - 277

Cunningham M R (1986) Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty *Journal of Personality and Social Psychology*, 50, 925 - 935

Cutler B L & Penrod S D (1988a) Context reinstatement and eyewitness identification In G M Davies & D M Thomson (eds.) *Memory in Context: Context in Memory* John Wiley & Sons, Chichester

Cutler B L & Penrod S D (1988b) Improving the reliability of eyewitness identification: Lineup construction and presentation *Journal of Applied Psychology*, 73, 281 - 290

Cutler B L & Penrod S D (1989) Moderators of the confidence-accuracy correlation in face recognition: The role of information processing and base-rates Applied Cognitive Psychology

Cutler B L, Penrod S D & Martens T K (1987a) Improving the reliability of eyewitness identification: Putting context into context *Journal of Applied Psychology*, 72, 629 - 637

Cutler B L, Penrod S D & Martens T K (1987b) The reliability of eyewitness identification: The role of system and estimator variables *Law and Human Behavior*, 11, 233 - 258

Cutler B L, Penrod S D, O'Rourke T E & Martens T K (1986) Unconfounding effects of contextual cues on eyewitness identification accuracy *Social Behaviour*, 1, 113 - 134

Darley J M & Gross P H (1983) A hypothesis-confirming bias in labelling effects *Journal of Personality and Social Psychology*, 44, 20 - 33

Darwin C (1904) *The Expression of Emotion in Man and Animals* John Murray, London

Davies G M (1981) Face Recall Systems In G Davies, H Ellis & J Shepherd (eds.) *Perceiving and Remembering Faces*, Academic Press: London

Davies G M (1983) Forensic face recall: The role of visual and verbal information In S M A Lloyd-Bostock & B R Clifford (eds.) *Evaluating Eyewitness Evidence: Recent Psychological Research and New Perspectives* John Wiley & Sons, Chichester

Davies G M, Ellis H D & Shepherd J W (1981) *Perceiving and Remembering Faces* Academic Press, London

Davies G M, Flin R H & Baxter J S (1986) The child witness *The Howard Journal* 25(2) 81 - 99

Davies G M & Milne A (1985) Recognizing faces in and out of context *Current Psychological Research*, 2, 235 - 246

Davies G M, Shepherd J W & Ellis H D (1979) Effects of interpolated mugshot exposure on accuracy of eyewitness identification *Journal of Applied Psychology*, 64, 232 - 237

Davies G M & Thomson D M (1988) *Memory in Context: Context in Memory* John Wiley & Sons: Chichester

Deaux K (1985) Sex and gender *Annual Review of Psychology*, 36, 49 - 81

Deaux K and Lewis L L (1984) Structure of gender stereotypes: Interrelationships among components and gender label *Journal of Personality and Social Psychology*, 46, 991 - 1004

Deffenbacher K & Horney J (1981) Psycho-legal aspects of face identification In G Davies, H Ellis & J Shepherd (eds.) *Perceiving and Remembering Faces*, Academic Press, London

Deffenbacher K A (1986) On the memorability of the human face In H D Ellis, M A Jeeves, F Newcombe & A Young (eds.) *Aspects of Face Processing* Martinus Nijhoff, Dordrecht

Deffenbacher K A (1991) A maturing of research on the behaviour of eyewitnesses *Applied Cognitive Psychology*, 5, 377 - 402

Deffenbacher K A, Brown E L & Sturgill W (1978) Some predictors of eyewitness memory accuracy In M M Gruneberg, P E Morris & R N Sykes (eds.) *Practical Aspects of Memory* Academic Press, London

Dent H R & Stephenson G M (1979) An experimental study of the effectiveness of different techniques of questioning child witnesses *British Journal of Social and Clinical Psychology*, 18, 41 - 51

Devine P G (1989a) Stereotypes and prejudice: Their automatic and controlled components *Journal of Personality and Social Psychology*, 56, 5 - 18

Devine P G (1989b) Automatic and controlled processes in prejudice: The role of stereotypes and personal beliefs In A R Pratkanis, S J Breckler & A G Greenwald (eds.) *Attitude, Structure and Function*, Lawrence Erlbaum Associates: Hillsdale, New Jersey

Dickens C (1832/1970) *Sketches by Boz I* Edito-Services S.A., Geneva

Dion K L (1979) Intergroup conflict and intergrouop cohesiveness In W G Austin & S Worchel (eds.) *The Social Psychology of Intergroup Relations* Brooks/Cole, Monterey, California

Dion K, Berscheid E, & Walster E (1972) "What is beautiful is good" *Journal of Personality and Social Psychology*, 24, 285 - 290

Dipboye R L, Fromkin H L & Wiback K (1975) Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant résumés *Journal of Applied Psychology*, 60, 39 - 43

Dooling D J & Christiaansen R E (1977) Episodic and semantic aspects of memory for prose *Journal of Experimental Psychology: Human Learning and Memory*, 3, 428 - 436

Duncan B L (1976) Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks *Journal of Personality and Social Psychology*, 34, 590 - 598

Ellis H D (1975) Recognising faces *British Journal of Psychology*, 66, 409 - 426

Ellis H D (1981) Theoretical aspects of face recognition In G Davies, H Ellis & J Shepherd (eds.) *Perceiving and Remembering Faces*, Academic Press, London

Ellis H D, Davies G M & Shepherd J W (1977) Experimental studies of face identification *National Journal of Criminal Defense*, 3, 219 - 234

Ellis H D, Shepherd J W & Davies G M (1980) The deterioration of verbal descriptions of faces over different delay intervals *Journal of Police Science and Administration*, 8, 101 - 106

Epps P & Parnell R W (1952) Physique and temperament of women delinquents compared with women undergraduates *British Journal of Medical Psychology*, 25, 249 - 255

Eysenck H J (1977) *Crime and Personality* Paladin, Hertfordshire

Eysenck M W (1984) *A Handbook of Cognitive Psychology* Lawrence Erlbaum Associates: London

Farnsworth P R (1934) Examinations in familiar and unfamiliar surroundings *Journal of Social Psychology*, 5, 128 - 129

Farrington D P (1981) Psychology and police interrogation *British Journal of Law and Society*, 8(1), 97 - 107

Faulkner W (1948/82) *Intruder in the Dust* Penguin, Aylesbury, Buckinghamshire

Fielder K, Hemmter U and Hofmann C (1984) On the origin of illusory correlation *European Journal of Social Psychology*, 14(2), 191 - 201

Firth A (1975) Interrogation *Police Review* 28th November

Fisher G & Cox R (1975) Recognizing human faces *Applied Ergonomics*, 6, 104 - 109

Fisher R P, Geiselman R E & Amador M (1989) Field test of the cognitive interviw: enhancing the recollection of actual victims and witnesses of crime *Journal of Applied Psychology*, 74, 722 - 727

Fisher R P, Geiselman R E & Raymond D S (1987) Critical analysis of the police interview *Journal of Police Science and Administration*, 15, 177 - 185

Fisher, R P, Geiselman R E, Raymond D S, Jurkevich L M & Warhafting M L (1987) Enhancing enhanced eyewitness memory: Refining the cognitive interview *Journal of Police Science and Administration*, 15(4), 291 - 297

Fishman J A (1956) An examination of the process and function of social stereotyping *Journal of Social Psychology*, 43, 26 - 64

Flammang C J (1972) Interviewing child victims of sex offenders *Police*, 16(6), 23 - 28

Flin R H (1980) Age effects in children's memory for unfamiliar faces *Developmental Psychology*, 16, 373 - 374

Flin R H (1988) The psychological evidence *New Law Journal* Aug. 26

Flin R H, Davies G M & Stevenson Y (1987) Children as witnesses: Psychological aspects of the English and Scottish systems *Medicine and the Law*, 6, 275 - 291

Flin R H & Shepherd J W (1986) Tall stories: Eyewitnesses' ability to estimate height and weight characteristics *Human Learning*, 5, 29 - 38

Flin R H, Stevenson Y & Davies G M (1989) Children's knowledge of court proceedings *British Journal of Psychology*, 80(3), 285 - 297

Flin R H & Tarrant M (1989) Children in the witness box *Social Work Today* Feb. 2

Friedemann V M & Morgan M K (1985) *Interviewing Sexual Abuse Victims Using Anatomical Dolls - The Professional's Guidebook* ed. L Anderson-Inman, Shamrock Press, Eugene, Oregon US

Friedman M, Reed S & Carterette E (1971) Feature saliency and recognition memory for schematic faces *Perception and Psychophysics*, 10, 47 - 50

Futoran G C & Wyer R S (1986) The effects of trait and gender stereotypes on occupational suitability judgments and the recall of judgment-relevant information *Journal of Experimental Social Psychology*, 22, 475 - 503

Garcia L T & Griffitt W (1978) The impact of testimonial evidence as a function of witness characteristics *Bulletin of the Psychonomic Society*, 11(1), 37 - 40

Gaskill P C, Fenton N and Porter J P (1927) Judging the intelligence of boys from their photographs *Journal of Applied Psychology*, 11, 394 - 403

Geiselman R E (1988) Improving memory through mental reinstatement of context In G M Davies & D M Thomson (eds.) *Memory in Context: Context in Memory*, John Wiley & Sons: Chichester

Geiselman R E, Fisher R P, Firstenberg I, Hutton L A, Sullivan S J, Avetissian I V & Prosk A L (1984) Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview *Journal of Police Science and Administration*, 12(1), 74 - 80

Geiselman R E, Fisher R P, MacKinnon D P & Holland H L (1985) Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis *Journal of Applied Psychology*, 70(2), 401 - 412

Geiselman R E, Fisher R P, MacKinnon D P & Holland H L (1986) Enhancement of eyewitness memory with the cognitive interview *American Journal of Psychology*, 99, 385 - 401

Geiselman R E & Padilla J (1988) Interviewing child witnesses with the cognitive interview *Journal of Police Science and Administration*, 16, 236 - 242

Geiselman R E, Saywitz K J & Bornstein G K (1992) Effects of cognitive questioning techniques on children's recall performance In G Goodman & B Bottoms (eds.) *Understanding and Improving Children's Testimony: Developmental, Clinical and Legal Isues*

George R (1992) The cognitive interview - does it work? Paper presented at the Third European Conference of Law and Psychology, 16th - 19th September, Oxford, UK

George R & Clifford B (1991) A field comparison of the cognitive interview and conversation management Paper presented at the British Psychological Society Annual Conference, Bournemouth, UK.

Gilbert G M (1951) Stereotype persistence and change among college students *Journal of Abnormal and Social Psychology*, 46, 245 - 254

Glueck S & Glueck G (1950) *Unravelling Juvenile Delinquency* Commonwealth Fund, New York

Godden D R & Baddeley A D (1975) Context-dependent memory in two natural environments: On land and underwater *British Journal of Psychology*, 66, 325 - 332

Goldstein A & Mackenberg E (1966) Recognition of human faces from isolated features: a developmental study *Psychonomic Science*, 6, 149 - 150

Goldstein A G, Stephenson B & Chance J (1977) Face recognition: Distribution of false alarms *Bulletin of the Psychonomic Society*, 9(6), 416 - 418

Goodsall J E (1974) The professional interviewer *Police Review* 26th April

Gorchinski S A (1973) Body weight and the perception of the weight of others *Dissertation Abstracts International*, 33, 6059B

Grant A (1987) Videotaping police questioning: A Canadian experiment *Criminal Law Review*, 375 - 385

Grant P R & Holmes J G (1981) The integration of implicit personality theory schemas and stereotype images *Social Psychology Quarterly*, 44, 107 - 115

Green E (1976) *Psychology for Law Enforcement* John Wiley & Sons, New York

Gudjonsson G H (1983) Suggestibility, intelligence, memory, recall and personality: An experimental study *British Journal of Psychiatry*, 142, 35 - 37

Gudjonsson G H (1986) The relationship between interrogative suggestibility and acquiescence: Empirical findings and theoretical implications *Personality and Individual Differences*, 7(2), 195 - 199

Gudjonsson G H & Lister S (1984) Interrogative suggestibility and its relationship with self-esteem and control *Journal of the Forensic Science Society*, 24, 99 - 110

Gurnee H (1934) An analysis of the perception of intelligence in the face *Journal of Social Psychology*, 5, 82 - 90

Hamilton D L (1978) Who is responsible? Towards a social psychology of pesponsibility attribution *Social Psychology*, 41, 316 - 328

Hamilton D L (1979) A cognitive-attributional analysis of stereotyping *Advances in Experimental Social Psychology*, 12, 53 - 85

Hamilton D L (1981a) Illusory correlation and stereotyping In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Hamilton D L (1981b) Stereotyping and intergroup behaviour: Some thoughts on the cognitive approach In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Hamilton D L & Gifford R K (1976) Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgements *Journal of Experimental and Social Psychology*, 12, 392 - 407

Hamilton D L & Rose T L (1980) Illusory correlation and the maintenance of stereotypic beliefs *Journal of Personality and Social Psychology*, 39(5), 832 - 845

Hanawalt N G (1944) The role of the upper and lower parts of the face as a basis for judging facial expressions: II. In posed and "candid" pictures *Journal of General Psychology*, 31, 23 - 36

Hashtroudi S, Mutter S A, Cole E A and Green S K (1984) Schema-consistent and schema-inconsistent information: Processing demands *Personality and Social Psychology Bulletin*, 10(2), 269 - 278

Hastie R & Kumar P A (1979) Person memory: Personality traits as organising principles in memory for behaviours *Journal of Personality and Social Psychology*, 37, 25 - 38

Hastie R & Park B (1986) The relationship between memory and judgment depends on whether the judgment task is memory-based on on-line *Psychological Review*, 93, 258 - 268

Hastie R, Park B and Weber R (1984) Social memory In R S Wyer and T K Srull *Handbook of Social Cognition*, Pt 2, Lawrence Erlbaum Associates: London

Heaton-Armstrong A (1985) Police officers' notebooks *Criminal Law Review* 781 - 784

Heller J (1962/88) *Catch-22* Corgi Books, Aylesbury, Buckinghamshire

Helson H (1964) *Adaptation-Level Theory* Harper & Rowe, New York

Hemsley G D & Marmurek H H C (1982) Person memory: The processing of consistent and inconsistent person information *Personality and Social Psychology Bulletin*, 8, 433 - 438

Heneman H G, III (1977) Impact of test information and applicant sex *Journal of Applied Psychology*, 62, 524 - 526

Hewstone M (1989) *Causal Attribution* Basil Blackwell Inc., Cambridge, Mass.

Higgins E T & Bargh J A (1987) Social cognition and social perception *Annual Review of Psychology*, 38, 369 - 425

Higgins E T & Rholes W S (1978) "Saying is believing": Effects of message modification on memory and liking for the person described *Journal of Experimental Social Psychology*, 14, 363 - 378

Hilgard E R & Loftus E F (1979) Effective interrogation of the eyewitness *The International Journal of Clinical and Experimental Hypnosis*, 27(4), 342 - 357

Hirschberg N, Jones L E T & Haggerty M (1978) What's in a face: Individual differences in face perception *Journal of Research in Personality*, 12, 488 - 499

Hochberg J & Galper R E (1967) Recognition of faces: I. An exploratory study *Psychonomic Science*, 9(12), 619 - 620

Hollin C R (1981) Nature of the witnessed incident and status of interviewer as variables influencing eyewitness recall *British Journal of Social Psychology*, 20, 295 - 296

Hollingworth L S (1935) The comparitive beauty of the faces of highly intelligent adolescents *Journal of Genetic Psychology*, 47, 268 - 281

Homa D, Haver B & Schwartz T (1976) Perceptibility of schematic face stimuli: evidence for a perceptual Gestalt *Memory and Cognition*, 4, 176 - 185

Horsch H M & Cooper D S (1982) Victimisation as a determinant of eyewitness accuracy *Journal of Applied Psychology*, 67(5), 649 - 652

Howard J W & Rothbart M (1980) Social categorization and memory for in-grouop and out-group behavior *Journal of Personality and Social Psychology*, 38, 301 - 310

Howe M J A (1983) *Introduction to the Psychology of Memory* Harper and Rowe, New York

Howells T H (1938) A study of ability to recognise faces *Journal of Abnormal and Social Psychology*, 33, 124 - 127

Hurwitz D, Wiggins N & Jones L (1975) Semantic differential for facial attributions: the face differential *Bulletin of the Psychonomic Society*, 6, 370 - 372

Irving B & Hilgendorf L (1980) *Police Interrogation: The Psychological Approach* Royal Commission on Criminal Procedure Study Crown Publishers Inc., New York

Johansson G (1973) Visual perception of biological motion and a model for its analysis *Perception and Psychophysics*, 14, 201 - 211

Johnson M K (1983) A multiple-entry, modular memory system In G H Bower (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 17) Academic Press, New York

Johnson M K (1985) The origin of memories In P C Kendall (ed.) *Advances in Cognitive Behavioural Research and Therapy*, 4, 1 - 27 Academic Press, New York

Johnson M K (1988) Discriminating the origin of information In T F Oltmanns & B A Maher (eds.) *Delusional Beliefs*, John Wiley and Sons, New York

Johnson M K & Foley M A (1984) Differentiating fact from fantsy: The reliability of children's memory *Journal of Social Issues*, 40, 33 - 50

Johnson M K, Foley H J, Raye C L & Foley M A (1981) Cognitive operations and decision bias in reality monitoring *American Journal of Psychology*, March, 94(1), 37 - 64

Johnson M K, Kahan T L & Raye C L (1984) Dreams and reality-monitoring *Journal of Experimental Psychology: General*, 113, 329 - 343

Johnson M K & Raye C L (1981) Reality monitoring *Psychological Review*, 88, 67 - 85

Johnson M K, Raye C L, Foley M A & Kim J (1982) Pictures and images: Spatial and temporal information compared *Bulletin of the Psychonomic Society*, 14, 23 - 26

Johnson M K, Raye C L, Hasher, L & Chromiak W (1979a) Are there developmental differences in reality-monitoring? *Journal of Experimental Child Psychology*, 27, 120 - 128

Johnson M K, Raye C L, Wang A Y & Taylor T H (1979b) Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences *Journal of Experimental Psychology: Human Learning and Memory*, 5, 229 - 240

Johnson M K & Suengas A G (1989) Reality monitoring judgments of other people's memories? *Bulletin of the Psychonomic Society*, 27(2), 107 - 110

Johnson M K, Taylor T H & Raye C L (1977) Fact and fantasy: The effects of internally-generated events on the apparent frequency of externally generated events *Memory and Cognition*, 5, 116 - 122

Jussim L, Coleman L M & Lerch L (1987) The nature of stereotypes: A comparison and integration of three theories *Journal of Personality and Social Psychology*, 52(3), 536 - 546

Kahan T L & Johnson M K (1990) Memory for seen and imagined rotations of alphanumeric characters *Journal of Mental Imagery*, 14(3), 119 - 129

Kahneman D, Slovic P & Tversky A (1982) *Judgment under Uncertainty: Heuristics and Biases* Cambridge University Press, New York

Kaplan J M (1990) Children don't always tell the truth *Journal of Forensic Science*, 35(3), 661 - 667

Karlins M, Coffman T L & Walters G (1969) On the fading of social stereotypes: Studies in three generations of college students *Journal of Personality and Social Psychology*, 13, 1 - 16

Katz D & Braly K W (1933) Racial stereotypes of one hundred college students *Journal of Abnormal and Social Psychology*, 28, 280 - 290

Keating C F (1985) Gender and the physiognomy of dominance and attractiveness *Social Psychology Quarterly*, 48, 61 - 70

Kelley H H (1971) *Attribution in Social Interaction* General Learning Press, Morristown, New Jersey

Kesey K (1986/87) *Demon Box* Methuen, London

Kiernan V G (1972) *Lords of Humankind: European Attitudes to the Outside World in the Imperial Age* Penguin, Hammondsworth, Middlesex

Kite M E & Deaux K (1987) Gender belief systems: Homosexuality and the implicit inversion theory *Psychology of Women Quarterly*, 11(1), 83 - 96

Klatzky R L, Martin G L & Kane R A (1982a) Influence of social-category activation on processing of visual information *Social Cognition*, 1, 95 - 109

Klatzky R L, Martin G L & Kane R A (1982b) Semantic interpretation effects on memory for faces *Memory and Cognition*, 10, 195 - 206

Koblinsky S A & Cruse D F (1981) The role of frameworks in children's retention of sex-related story content *Journal of Experimental Child Psychology*, 31(2), 321 - 331

Koepp D P, Lubbers H D, Lucero R C, Mankiewicz N L & Mason D C (1981) To interview a fire witness *Fire/Arson Research Paper*, November 2-20, 1981 United States

Köhnken G, Finger M, Nitschke N & Höfer E (1992a) Does a cognitive interview interfere with a subsequent statement validity analysis? Paper presented at the Conference of the American Psychology-Law Society in San Diego

Köhnken G, Schimossek E, Aschermann E & Höfer E (1991) The cognitive interview and statement validity analysis Paper presented at the International Conference on Memory, Lancaster, UK

Kozlowski L T & Cutting J E (1977) Recognising the sex of a walker from a dynamic point-light display *Perception and Psychophysics*, 21, 575 - 580

Kuehn L L (1974) Looking down a gun barrel: Person perception and violent crime *Perceptual and Motor Skills*, 39, 1159 - 1164

Langdell T (1978) Recognition of faces: An approach to the study of autism *Journal of Child Psychology and Psychiatry*, 19, 255 - 268

Laughery K R, Fessler P K, Lenorovitz D R & Yoblick D A (1974) Time delay and similarity effects in facial recognition *Journal of Applied Psychology*, 59, 490 - 496

Leippe M R, Manion A P & Romanczyk A (1992) Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology*, 63(2), 181 - 197

Lepper M R, Ross L & Lau R R (1986) Persistence of inaccurate beliefs about the self: Perseverance effects in the classroom *Journal of Personality and Social Psychology*, 50(3), 482 - 491

Lieppe M R, Wells G & Ostrom T M (1978) Crime seriousness as a determinant of accuracy in eyewitness identification *Journal of Applied Psychology*, 63, 345 - 351

Lindsay D S & Johnson M K (1987) Reality monitoring and suggestibility: Children's ability to discriminate among memories from different sources In S J Ceci, M P Toglia and D F Ross (eds.) *Children's Eyewitness Testimony* Springer - Verlag, New York.

Lindsay D S & Johnson M K (1989a) The eyewitness suggestibility effect & memory for source *Memory and Cognition*, 17(3), 349 - 358

Lindsay D S & Johnson M K (1989b) The reversed eyewitness suggestibility effect *Bulletin of the Psychonomic Society*, 27(2), 111 - 113

Lindsay D S & Johnson M K (1991) Recognition memory and source-monitoring *Bulletin of the Psychonomic Society*, 29(3), 203 - 205

Lindsay D S (1990) Misleading suggestions can impair eyewitnesses' ability to remember event details *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 1077 - 1083

Lindsay R C, Wallbridge H & Drennan D (1987) Do the clothes make the man? An exploration of the effect of lineup attire on eyewitness identification accuracy *Canadian Journal of Behavioral Science Special Issue: Forensic Psychology*, 19(4), 463 - 478

Lingle J H, Geva N, Ostrom T M, Leippe M R & Baumgartner M H (1979) Thematic effects of person judgments on impression organization *Journal of Personality and Social Psychology*, 37(5), 674 - 687

Litman G K, Powell G E & Stewart R A (1983) Fine-grained stereotyping and the structure of social cognition *Journal of Social Psychology*, 120(1), 45 - 56

Locksley A, Borgida E, Brekke N & Hepburn C (1980) Sex stereotypes and social judgment *Journal of Personality and Social Psychology*, 39, 821 - 831

Locksley A, Hepburn C and Ortiz V (1982) Social stereotypes and judgments of individuals: An instance of the base-rate fallacy *Journal of Experimental Social Psychology*, 18, 23 - 42

Locksley A, Ortiz V & Hepburn C (1980) Social categorization and discriminatory behaviour: Extinguishing the minimal intergroup discrimination effect *Journal of Personality and Social Psychology*, 39(5), 773 - 783

Loftus E F (1975) Leading questions and eyewitness report *Cognitive Psychology*, 7, 560 - 572

Loftus E F (1977) Shifting human color memory *Memory and Cognition*, 5, 696 - 699

Loftus E F & Hoffman H G (1989) Misinformation and memory: The creation of new memories *Journal of Experimental Psychology: General*, 118(1), 100 - 104

Loftus E F, Korf N L & Schooler J W (1989) Misguided memories: Sincere distortions of reality In J C Yuille (ed.) *Credibility Assessment* 155 - 73. Kluwer Academic Publishers: Dordrecht

Loftus E F, Loftus G R & Messo J (1987) Some facts about "weapon focus" *Law and Human Behavior*, 11, 55 - 62

Loftus E F & Marburger W (1983) Since the eruption of Mt St Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events *Memory and Cognition*, 11(2) 114 - 120

Loftus E F, Miller D G & Burns H J (1978) Semantic integration of verbal information into visual memory *Journal of Experimental Psychology: Human Learning and Memory*, 4, 19 - 31

Loftus E F & Palmer J C (1974) Reconstruction of automobile destruction: An example of the interaction between language and memory *Journal of Verbal Learning and Verbal Behavior*, 13, 585 - 564

Loftus E F, Schooler J W & Wagenaar W A (1985) The fate of memory: Comment of McCloskey and Zaragoza *Journal of Experimental Psychology: General*, 114, 375 - 380

Loftus E F & Zanni G (1975) Eyewitness testimony: The influence of the wording of a question *Bulletin of the Psychonomic Society*, 2, 86 - 88

Lui L & Brewer M B (1983) Recognition accuracy as evidence of category-consistency effects in person memory *Social Cognition*, 2(2), 89 - 107

Luria S M and Strauss M S (1978) Comparison of eye movements over faces in photographic positives and negatives *Peception*, 7, 349 - 358

MacLeod M D & Frowley J N (1991) A psychological analysis of the evidence presented against Robert Campbell For BBC TV, *Rough Justice*

MacLeod M D, Carson L & Prescott R G W (in press) Ethnic minorities and the criminal justice system In M D MacLeod & C N Macrae (eds.) *Stereotypes in the Criminal Justice System*, Falmer Press, London

MacLeod M D, Frowley J N & Shepherd J W (in press) Whole body information: Its relevance to eyewitnesses In D F Ross, J D Reid & M P Toglia (eds) *Adult Eyewitness Memory: Current Trends and Developments* Cambridge University Press, New York

Macrae C N (in press) Cognitive misers in the courtroom In M D MacLeod & C N Macrae (eds.) *Stereotypes in the Criminal Justice System*, Falmer Press, London

Macrae C N & Shepherd J W (1989a) Do criminal stereotypes mediate juridic judgements? *British Journal of Social Psychology*, 28, 189 - 191

Macrae C N & Shepherd J W (1989b) Stereotypes and social judgments *British Journal of Social Psychology*, 28, 319 - 325

Macrae C N & Shepherd J W (1991) Categorical effects on attributional inferences: A response-time analysis *British Journal of Social Psychology*, 30, 235 - 245

Malpass R S & Devine P G (1981a) Eyewitness identification: lineup instructions and the absence of the offender *Journal of Applied Psychology*, 66, 482 - 489.

Malpass R S & Devine P G (1981b) Guided memory in eyewitness identification *Journal of Applied Psychology*, 66, 33 - 50

Malpass R S & Devine P G (1983) Measuring the fairness of eyewitness identification lineups In S M A Lloyd-Bostock & B R Clifford (eds.) *Evaluating Eyewitness Evidence* John Wiley & Sons, Chichester

Malpass R S & Devine P G (1984) Research on suggestion in lineups and photospreads In G Wells & E Loftus (eds) *Eyewitness Testimony* Cambridge University Press, New York

Malpass R S & Kravitz J (1969) Recognition for faces of own and other race *Journal of Personality and Social Psychology*, 13(4), 330 - 334

Malpass R S (1981) Training in face recognition In G M Davies, H D Ellis & J W Shepherd (eds.) *Perceiving and Remembering Faces*, Academic Press, London

Manzanero A L & Diges M (1992) Effects of preparation on internal and external memories Paper presented at Third European Conference of Law & Psychology, 16 - 19 September 1992

Marquis K H, Marshall J & Oskamp S (1972) Testimony validity as a function of question form, atmosphere, and item difficulty Journal of Applied Psychology, 2, 167 - 186

Marshall J (1966) *Law and Psychology in Conflict* Bobbs-Merrill, New York

McArthur L Z (1982) Judging a book by its cover: A cognitive analysis of the relationship between physical appearance and stereotyping In A Hastorf & A Isen (eds.) *Cognitive Social Psychology*, Elsevier, New York

McArthur L Z & Apatow K (1983 - 84) Impressions of baby-faced adults *Social Cognition*, 2, 315 - 342

McArthur L Z & Apatow K (1983 - 84) Impressions of baby-faced adults *Social Cognition*, 2, 315 - 342

McCauley C, Stitt C L & Segal M (1980) Stereotyping: From Prejudice to Prediction *Psychology Bulletin*, 87(1), 195 - 208

McCloskey M & Zaragoza M (1985) Misleading Postevent Information and Memory for Events: Arguments and Evidence Against the Memory Impairment Hypothesis *Journal of Experimental Psychology: General*, 114(1), 1 - 16

McEwan I (1975/76) *First Love, Last Rites*, Picador, London

McKelvie S (1976) The role of eyes and mouth in recognition memory for faces *American Journal of Psychology*, 89, 311 - 323

McKelvie S J (1978) Sex differences in facial memory In *Practical Aspects of Memory*, M M Gruneberg, P E Morris & R N Sykes (eds.) Academic Press, London

Memon A & Bull R (1991) The cognitive interview, its origins, empirical support, evaluation and practical implications *Journal of Community and Applied Social Psychology*, 1, 291 - 307

Memon A, Cronin O, Eaves R, Bull R & Küpper B (1992) The cognitive interview and child witnesses Paper presented at ther NATO Advanced Study Institute: The child witness in context: cognitive, social and legal perspectives, Lucca, Italy

Memon A & Köhnken G (1992) Helping witnesses to remember more: The cognitive interview Paper presented at the Third Annual Conference of Law and Psychology, 16th - 19th September, Oxford, UK

Mende L, MacKinnon D P & Geiselman R E (1986) Memory for licence plates as a function of exposure time *Journal of Police Science and Administration*, 15(1), 68 - 71

Metcalfe J (1990) Composite holographic associative recall model (CHARM) and blended memories in eyewitness testimony *Journal of Experimental Psychology: General*, 119(2), 145 - 160

Miller D T, Turnbull W & McFarland C (1990) Counterfactual thinking and social perception: Thinking about what might have been *Advances in Experimental Social Psychology*, 23, 305 - 331

Milord J T (1978) Aesthetic aspects of faces: A (somewhat) phenomenological analysis using multi-dimensional scaling methods *Journal of Personality and Social Psychology*, 36, 205 - 216

Mortimer A (1991) "A case of..." Patterns in police officers' perceptions of offences Paper presented at the BPS Division of Criminological and Legal Psychology Conference, University of Kent, Canterbury, 3rd - 5th January

Moston S (1988) Peer Social Support and Children's Eyewitness Memory Paper presented at BPS Social Psychology Section Annual Conference, University of Kent, Canterbury 23rd September

Moston S (1989) If children's testimony is reliable, why don't we rely on it? Paper presented at BPS Annual Conference, University of St Andrews, 1st April

Nakdimen K A (1984) The physiognomic basis of sexual stereotyping *American Journal of Psychiatry*, 141, 399 - 503

Nash H (1969) Rocognition of body-surface regions *Genetic Psychology Monographs*, 79, 297 - 340

Newlands P (1993) Evaluating the cognitive interview Paper presented at PSI - PAG Conference, June 11th - 13th, University of Coleraine, Northern Ireland

Nisbett R E & Ross L (1980) *Human Inference: Strategies and Shortcomings of Social Judgment* Prentice-Hall, Englewood Cliffs, New Jersey

Nobbs D (1977) *The Return of Reginald Perrin* Penguin, Middlesex

Oakes P J & Turner J C (1980) Social categorisation and intergroup behaviour: Does minimal group discrimination make social identity more postive? *European Journal of Social Psychology*, 10, 295 - 302

Orne M T, Soskis D A, Dinges D F & Orne E C (1984) Hypnotically Induced Testimony In G Wells and E F Loftus (eds.) *Eyewitness Testimony* Cambridge University Press, New York

O'Sullivan C S & Durso F T (1984) Effect of schema-incongruent information on memory for stereotypical attributes *Journal of Personality and Social Psychology*, 47(1), 53 - 70

Palumbo P A (1975) Effective Interviewing - More Than a Mere Process of Questioning *Law and Order*, 23(9), 18, 20-23

Park B & Rothbart M (1982) Perception of out-group homogeneity and levels of social categorization: Memory for the subordinate attributes of in-group and out-group members *Journal of Personality and Social Psychology*, 42(6), 1051 - 1068

Patten B M (1990) The history of the memory arts *Neurology*, 40, 346 - 352

Patterson K E & Baddeley A D (1977) When face recognition fails *Journal of Experimental Psychology: Human Learning and Memory*, 3(4), 406 - 417

Pennington N & Hastie R (1986) Evidence evaluation in complex decision making *Journal of Personality and Social Psychology*, 51, 242 - 258

Perky C W (1910) An experimental study of imagination *American Journal of Psychology*, 21, 422 - 452

Perrett D I, Mistlin A J & Chitty A J (1987) Visual neurones responsive to faces *Trends in Neurosciences*, 10(9)

Perry D J & Chapman A J (1992) Applying the cognitive interview procedure to road accident witnesses Paper presented at the Annual Conference of the British Psychological Society

Pettigrew T F (1981) Extending the stereotype concept In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Petzold P (1986) The influence of category representation in social judgments In F Klix & H Hagendorf (eds.) *Human Memory and Cognitive Capacities: Mechanisms and Performance*, Elsevier, Amsterdam, Netherlands

Posner M I & Snyder C R (1975) Attention and cognitive control In R L Solso (ed.) *Information Processing and Cognition: The Loyola Symposium* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Putnam W H (1979) Hypnosis and distortion in eyewitness memory *International Journal of Clinical and Experimental Hypnosis*, 27, 437 - 448

Rabbie J M & Horwitz M (1969) Arousal of ingroup-outgroup biasby a chance win or loss *Journal of Personality and Social Psychology*, 13, 269 - 277

Rees L (1973) Constitutional factors and abnormal behaviour In H J Eysenck (ed) *Handbook of Abnormal Psychology* Pitman Medical, London

Reiser M R (1974) Hypnosis as an aid in a homicide investigation *The American Journal of Clinical Hypnosis*, 17(2), 84 - 87

Reiser M R (1976) *Handbook of Investigative Hypnosis* LEHI: Los Angeles

Rice S A (1926) "Stereotypes": A source of error in judging human character *Journal of Personnel Research*, 5, 267 - 276

Roberts J V (1985) The attention-memory relationship after forty years: A meta-analysis of the literature *Basic and Applied Social Psychology*, 6(3), 221 - 241

Roediger III H L & Payne D G (1983) Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories *Memory and Cognition*, 13(1), 1 - 7

Roediger III H L, Srinivas K & Waddill P (1989) How much does guessing influence recall? Comment on Erdelyi, Finks and Feigin-Pfau *Journal of Experimental Psychology - General*, 118(3), 255 - 257

Ross D F, Miller B S & Moran P B (1987) The child in the eyes of the jury: Assessing mock jurors' perceptions of the child witness In S J Ceci, M P Toglia and D F Ross (eds.) *Children's Eyewitness Testimony* Springer-Verlag, New York.

Ross L, Lepper M R, Strack F & Steinmetz J (1977) Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood *Journal of Personality and Social Psychology*, 35(11), 817 - 829

Ross M, McFarland C & Fletcher G (1981) The effect of attitude on the recall of personal histories *Journal of Personality and Social Psychology*, 40(4), 627 - 634

Rosselli M (1984) Police lose, but taping wins *Sunday Times* 1st July p 7

Rothbart M (1981) Memory processes and social beliefs In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Rothbart M, Evans M & Fulero S (1979) Recall for confirming events: Memory processes and the maintenance of social stereotypes *Journal of Experimental Psychology*, 15, 343 - 355

Roy D F (1991) Towards improved recall by an eyewitness - a review of the cognitive interview and other techniques Paper presented at the First Annual Conference Of The British Psychological Society Division of Criminological and Legal Psychology, 3rd - 5th January 1991, University of Kent, Canterbury.

Ruble D N & Stangor C (1986) Stalking the elusive schema: Insights from developmental and social psychological analyses of gender schemas *Social Cognition*, 4(2), 227 - 261

Ryckman R M, Robbins M A, Kaczor L M & Gold J A (1989) Male and female raters' stereotyping of male and female physiques *Personality and Social Psychology Bulletin*, 15(2) 244 - 251

Sagar H & Schofield J (1980) Racial and behavioural cues in black and white children's perceptions of ambiguously aggressive acts *Journal of Personality and Social Psychology*, 39, 590 - 598

Sanders G S & Chiu W (1988) Eyewitness errors in the free recall of actions *Journal of Applied Social Psychology*, 18(14), 1241 - 1259

Scheithauer E, Woratschek G and Tscherne W (1983) *Geschichte Osterreichs in Stichworten:V - Die Zeit der Demokratischen Republik Osterreich von 1918 bis 1934* Verlag Ferdinand Hirt, Vienna

Schmitz W & Plate M (1978) Police interrogation and offence facts - Results of an empirical study of reconstruction of an offence by interviews of witnesses *Krimimalistik*, 32(12), 541 - 546

Schneider D J (1973) Implicit personality theory: A review *Psychological Bulletin*, 79, 293 - 309

Schooler J W, Gerhard D & Loftus E F (1986) Qualities of the unreal *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12(2), 171 - 181

Schuh A J (1978) Effects of an early interruption and note-taking on listening accuracy and decision-making in the interview *Bulletin of the Psychonomic Society*, 12(3), 242 - 244

Schuh A J (1980) Verbal listening skill in the interview and personal characteristics of the listeners *Bulletin of the Psychonomic Society*, 15(2), 125 - 127

Schuh A J (1981) Establishing, maintaining and evaluating and interview training program *Bulletin of the Psychonomic Society*, 17(3), 143 - 146

Seamon J G, Stolz J A, Bass D H and Chatinover A I (1978) Recognition of facial features in immediate memory *Bulletin of the Psychonomic Society*, 12, 231 - 234

Secord P F & Bevan W (1956) Personalities in faces: III. A cross-cultural comparison of impressions of physiognomy and personality in faces *Journal of Social Psychology*, 43, 283 - 288

Secord P F & Muthard J E (1955) Personalities in faces: IV. A descriptive analysis of the perception of women's faces and the identification of some physiognomic determinants *Journal of Psychology*, 39, 269 - 278

Secord P F (1958) Facial features and inference processes in interpersonal perception In R Tagiuri & L Petrullo (eds.) *Person Perception and Interpersonal Behaviour*, Stanford University Press, Stanford

Secord P F (1959) Stereotyping and favourableness in the perception of negro faces *Journal of Abnormal and Social Psychology*, 59, 309 - 314

Secord P F, Bevan W & Dukes W F (1953) Occupational and physiognomic stereotypes in the perception of photographs *Journal of Social Psychology*, 37, 261 - 270

Secord P F, Bevan W & Katz B (1956) The negro stereotype and perceptual accentuation *Journal of Abnormal and Social Psychology*, 53, 78 - 83

Secord P F, Dukes W F & Bevan W (1954) Personalities in faces: I. An experiment in social perceiving *Genetic Psychology Monographs*, 49, 231 - 279

Sentis K P & Burnstein E (1979) Remembering schema-consistent information: Effects of a balance schema on recognition memory *Journal of Personality and Social Psychology*, 37(12), 2200 - 2211

Shapiro P N & Penrod S D (1986) Meta-analysis of face identification studies *Psychological Bulletin*, 100, 139 - 156

Sheikh J I, Hill R D & Yesavage J A (1986) Long-term efficacy of cognitive training for age-associated memory impairment: A six-month follow-up study *Developmental Neuropsychology*, 2(4), 413 - 421

Sheldon W H, Stevens S S & Tucker W B (1940) *The Varieties of Human Physique: An Introduction to Constitutional Psychology* Harper & Brothers, New York

Shepherd E (1985) Interviewing development: facing up to reality Paper presented at the NATO Advanced Study Institute on the Role of Psychology in the Selection and Training of Police, 7th - 15th May 1985 Skiathos, Greece

Shepherd E (1991) Resistance in interviews: The contribution of police perceptions and behaviour Paper Presented at the Division of Criminological and Legal Psychology Conference, 3rd - 5th January 1991 University of Kent at Canterbury.

Shepherd J W & Ellis H D (1973) The effect of attractiveness on recognition memory for faces *American Journal of Psychology*,86, 627 - 633

Shepherd J W, Ellis H D & Davies G M (1977) Perceiving and remembering faces Technical report to the Home Office under contract POL/73/1675/24/1

Shepherd J W, Ellis H D & Davies G M (1981) Studies of cue saliency In G M Davies, H D Ellis & J Shepherd (eds) *Perceiving and Remembering Faces* Aberdeen University Press, Aberdeen

Shepherd J W, Ellis H D & Davies G M (1982) *Identification Evidence: A Psychological Evaluation* Aberdeen University Press, Aberdeen

Sherif M (1966) *Group Conflict and Co-operation: Their Social Psychology* Routledge and Kegan-Paul, London

Sherman S J & Corty E (1984) Cognitive heuristics In R S Wyer & T K Srull (eds.) *Handbook of Social Cognition Vol. 1* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Sherman S J, Judd C M and Park B (1989) Social cognition *Annual Review of Psychology*, 40, 281 - 326

Shields J (1973) Heredity and psychological development In H J Eysenck (ed) *Handbook of Abnormal Psychology* London, Pitman Medical

Shirer W (1959/60) *The Rise and Fall of the Third Reich* Redwood Press, Wiltshire, UK

Shoemaker D J, South D R & Lowe J (1973) Facial stereotypes of deviants and judgements of guilt or innocence *Social Forces*, 51, 427 - 433

Sinha A K P & Upadhyaya U P (1960) Change and persistence in the stereotypes of university students towards different ethnic groups during the Sino-Indian border dispute *Journal of Social Psychology*, 52, 31 - 39

Smith S M (1979) Remembering in and out of context *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460 - 471, 5, 460 - 471

Smith V L & Ellsworth P C (1987) The social psychology of eyewitness accuracy: misleading questions and communicator expertise *Journal of Applied Psychology*, 72(2), 294 - 300

Smith S M, Glenberg A & Bjork R A (1978) Environmental context and human memory *Memory and Cognition*, 6, 342 - 353

Snyder M (1981) On the self-perpetuating nature of social stereotypes In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1981

Snyder M (1984) When belief creates reality *Advances in Experimental Social Psychology*, 18, 247 - 303

Snyder M & Uranowitz S W (1978a & b) Unpublished studies, cited in Snyder M (1981) *op cit.*

Snyder M & Uranowitz S W (1978c) Reconstructing the past: some cognitive consequences of person perception *Journal of Personality and Social Psychology*, 36(9), 941- 950

Sobell L C, Toneatto T, Sobell M B, Schuller R & Maxwell M (1990) A procedure for reducing errors in reports of life events *Journal of Psychosomatic Research*, 34(2), 163 - 170

Softley P, Brown D, Forde B, Mair G & Moxon D (1980) Police interrogation - an observational study in four police stations Royal Commission on Criminal Procedure HMSO, London

Sporer S L (1982) A brief history of the psychology of testimony *Current Psychological Reviews*, 2, 323 - 340.

Srull T K & Wyer R S (1989) Person memory and judgment *Psychological Review*, 96(1), 58 - 83

Srull T K (1984) Methodological techniques for the study of person memory and social cognition In R S Wyer & T K Srull (eds.) *Handbook of Social Cognition Vol 2* Lawrence Erlbaum Associates, New Jersey

Stangor C & Ruble D N (1989) Strength of expectancies and memory for social information: What we remember depends on how kmuch we know *Journal of Experimental Social Psychology*, 25, 18 - 35

Steller M (1989) Recent developments in statement analysis In J C Yuille (ed.) *Credibility Assessment* Kluwer Academic Publishers, London

Stephenson G M, Clark N K & Wade G S (1986) Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation *Journal of Personality and Social Psychology*, 50(6), 1113 - 1122

Stern W (1904) Wirklichkeitversuche *Beitrage zur Psychologie der Aussage*, 2(1), 1 - 31

Stevens D & Berliner L (1976) Special techniques for child witnesses In L G Schultz (ed.) *The Sexual Victimology of Youth* C C Thomas, London

Strand B Z (1970) Change of context and retroactive inhibition *Journal of Verbal Learning and Verbal Behavior*, 9, 202 - 206

Strongman K T & Hart C J (1968) Stereotyped reactions to body-build *Psychological Reports*, 23, 1175 - 1178

Suengas A G & Johnson M K (1985) Effects of rehearsal of perceived and imagined memories. Paper presented at the meeting of the Eastern Psychological Association, March

Sussman W G A (1992) Improving eyewitness recall using the source-monitoring technique Unpublished MA Thesis, University of St Andrews

Tajfel H (1959) Quantitative judgement in social perception *British Journal of Psychology*, 10, 16 - 29

Tajfel H (1957) Value and judgement of magnitude *Psychological Review*, 64, 192 - 204

Tajfel H (1969a) Cognitive aspects of prejudice *Journal of Social Issues*, 25, 79 - 97

Tajfel H (1969b) Social and cultural factors in perception In G Lindzey & E Aronson (eds) *The Handbook of Social Psychology, 2nd edition*, Addison-Wesley, Reading, Mass.

Sanders G S & Chiu W (1988) Eyewitness errors in the free recall of actions *Journal of Applied Social Psychology*, 18(14), 1241 - 1259

Scheithauer E, Woratschek G and Tscherne W (1983) *Geschichte Osterreichs in Stichworten:V - Die Zeit der Demokratischen Republik Osterreich von 1918 bis 1934* Verlag Ferdinand Hirt, Vienna

Schmitz W & Plate M (1978) Police interrogation and offence facts - Results of an empirical study of reconstruction of an offence by interviews of witnesses *Krimimalistik*, 32(12), 541 - 546

Schneider D J (1973) Implicit personality theory: A review *Psychological Bulletin*, 79, 293 - 309

Schooler J W, Gerhard D & Loftus E F (1986) Qualities of the unreal *Journal of Experimental Psychology: Learning, Memory and Cognition*, 12(2), 171 - 181

Schuh A J (1978) Effects of an early interruption and note-taking on listening accuracy and decision-making in the interview *Bulletin of the Psychonomic Society*, 12(3), 242 - 244

Schuh A J (1980) Verbal listening skill in the interview and personal characteristics of the listeners *Bulletin of the Psychonomic Society*, 15(2), 125 - 127

Schuh A J (1981) Establishing, maintaining and evaluating and interview training program *Bulletin of the Psychonomic Society*, 17(3), 143 - 146

Seamon J G, Stolz J A, Bass D H and Chatinover A I (1978) Recognition of facial features in immediate memory *Bulletin of the Psychonomic Society*, 12, 231 - 234

Secord P F & Bevan W (1956) Personalities in faces: III. A cross-cultural comparison of impressions of physiognomy and personality in faces *Journal of Social Psychology*, 43, 283 - 288

Secord P F & Muthard J E (1955) Personalities in faces: IV. A descriptive analysis of the perception of women's faces and the identification of some physiognomic determinants *Journal of Psychology*, 39, 269 - 278

Secord P F (1958) Facial features and inference processes in interpersonal perception In R Tagiuri & L Petrullo (eds.) *Person Perception and Interpersonal Behaviour*, Stanford University Press, Stanford

Secord P F (1959) Stereotyping and favourableness in the perception of negro faces *Journal of Abnormal and Social Psychology*, 59, 309 - 314

Secord P F, Bevan W & Dukes W F (1953) Occupational and physiognomic stereotypes in the perception of photographs *Journal of Social Psychology*, 37, 261 - 270

Secord P F, Bevan W & Katz B (1956) The negro stereotype and perceptual accentuation *Journal of Abnormal and Social Psychology*, 53, 78 - 83

Secord P F, Dukes W F & Bevan W (1954) Personalities in faces: I. An experiment in social perceiving *Genetic Psychology Monographs*, 49, 231 - 279

Sentis K P & Burnstein E (1979) Remembering schema-consistent information: Effects of a balance schema on recognition memory *Journal of Personality and Social Psychology*, 37(12), 2200 - 2211

Shapiro P N & Penrod S D (1986) Meta-analysis of face identification studies *Psychological Bulletin*, 100, 139 - 156

Sheikh J I, Hill R D & Yesavage J A (1986) Long-term efficacy of cognitive training for age-associated memory impairment: A six-month follow-up study *Developmental Neuropsychology*, 2(4), 413 - 421

Sheldon W H, Stevens S S & Tucker W B (1940) *The Varieties of Human Physique: An Introduction to Constitutional Psychology* Harper & Brothers, New York

Shepherd E (1985) Interviewing development: facing up to reality Paper presented at the NATO Advanced Study Institute on the Role of Psychology in the Selection and Training of Police, 7th - 15th May 1985 Skiathos, Greece

Shepherd E (1991) Resistance in interviews: The contribution of police perceptions and behaviour Paper Presented at the Division of Criminological and Legal Psychology Conference, 3rd - 5th January 1991 University of Kent at Canterbury.

Shepherd J W & Ellis H D (1973) The effect of attractiveness on recognition memory for faces *American Journal of Psychology*,86, 627 - 633

Shepherd J W, Ellis H D & Davies G M (1977) Perceiving and remembering faces Technical report to the Home Office under contract POL/73/1675/24/1

Shepherd J W, Ellis H D & Davies G M (1981) Studies of cue saliency In G M Davies, H D Ellis & J Shepherd (eds) *Perceiving and Remembering Faces* Aberdeen University Press, Aberdeen

Shepherd J W, Ellis H D & Davies G M (1982) *Identification Evidence: A Psychological Evaluation* Aberdeen University Press, Aberdeen

Sherif M (1966) *Group Conflict and Co-operation: Their Social Psychology* Routledge and Kegan-Paul, London

Sherman S J & Corty E (1984) Cognitive heuristics In R S Wyer & T K Srull (eds.) *Handbook of Social Cognition Vol. 1* Lawrence Erlbaum Associates, Hillsdale, New Jersey

Sherman S J, Judd C M and Park B (1989) Social cognition *Annual Review of Psychology*, 40, 281 - 326

Shields J (1973) Heredity and psychological development In H J Eysenck (ed) *Handbook of Abnormal Psychology* London, Pitman Medical

Shirer W (1959/60) *The Rise and Fall of the Third Reich* Redwood Press, Wiltshire, UK

Shoemaker D J, South D R & Lowe J (1973) Facial stereotypes of deviants and judgements of guilt or innocence *Social Forces*, 51, 427 - 433

Sinha A K P & Upadhyaya U P (1960) Change and persistence in the stereotypes of university students towards different ethnic groups during the Sino-Indian border dispute *Journal of Social Psychology*, 52, 31 - 39

Smith S M (1979) Remembering in and out of context *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460 - 471, 5, 460 - 471

Smith V L & Ellsworth P C (1987) The social psychology of eyewitness accuracy: misleading questions and communicator expertise *Journal of Applied Psychology*, 72(2), 294 - 300

Smith S M, Glenberg A & Bjork R A (1978) Environmental context and human memory *Memory and Cognition*, 6, 342 - 353

Snyder M (1981) On the self-perpetuating nature of social stereotypes In D L Hamilton (ed.) *Cognitive Processes in Stereotyping and Intergroup Behaviour* Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1981

Snyder M (1984) When belief creates reality *Advances in Experimental Social Psychology*, 18, 247 - 303

Snyder M & Uranowitz S W (1978a & b) Unpublished studies, cited in Snyder M (1981) *op cit.*

Snyder M & Uranowitz S W (1978c) Reconstructing the past: some cognitive consequences of person perception *Journal of Personality and Social Psychology*, 36(9), 941- 950

Sobell L C, Toneatto T, Sobell M B, Schuller R & Maxwell M (1990) A procedure for reducing errors in reports of life events *Journal of Psychosomatic Research*, 34(2), 163 - 170

Softley P, Brown D, Forde B, Mair G & Moxon D (1980) Police interrogation - an observational study in four police stations Royal Commission on Criminal Procedure HMSO, London

Sporer S L (1982) A brief history of the psychology of testimony *Current Psychological Reviews*, 2, 323 - 340.

Srull T K & Wyer R S (1989) Person memory and judgment *Psychological Review*, 96(1), 58 - 83

Srull T K (1984) Methodological techniques for the study of person memory and social cognition In R S Wyer & T K Srull (eds.) *Handbook of Social Cognition Vol 2* Lawrence Erlbaum Associates, New Jersey

Stangor C & Ruble D N (1989) Strength of expectancies and memory for social information: What we remember depends on how kmuch we know *Journal of Experimental Social Psychology*, 25, 18 - 35

Steller M (1989) Recent developments in statement analysis In J C Yuille (ed.) *Credibility Assessment* Kluwer Academic Publishers, London

Stephenson G M, Clark N K & Wade G S (1986) Meetings make evidence? An experimental study of collaborative and individual recall of a simulated police interrogation *Journal of Personality and Social Psychology*, 50(6), 1113 - 1122

Stern W (1904) Wirklichkeitversuche *Beitrage zur Psychologie der Aussage*, 2(1), 1 - 31

Stevens D & Berliner L (1976) Special techniques for child witnesses In L G Schultz (ed.) *The Sexual Victimology of Youth* C C Thomas, London

Strand B Z (1970) Change of context and retroactive inhibition *Journal of Verbal Learning and Verbal Behavior*, 9, 202 - 206

Strongman K T & Hart C J (1968) Stereotyped reactions to body-build *Psychological Reports*, 23, 1175 - 1178

Suengas A G & Johnson M K (1985) Effects of rehearsal of perceived and imagined memories. Paper presented at the meeting of the Eastern Psychological Association, March

Sussman W G A (1992) Improving eyewitness recall using the source-monitoring technique Unpublished MA Thesis, University of St Andrews

Tajfel H (1959) Quantitative judgement in social perception *British Journal of Psychology*, 10, 16 - 29

Tajfel H (1957) Value and judgement of magnitude *Psychological Review*, 64, 192 - 204

Tajfel H (1969a) Cognitive aspects of prejudice *Journal of Social Issues*, 25, 79 - 97

Tajfel H (1969b) Social and cultural factors in perception In G Lindzey & E Aronson (eds) *The Handbook of Social Psychology, 2nd edition*, Addison-Wesley, Reading, Mass.

Tajfel H (1970) Experiments in intergroup discrimination *Scientific American*, 223, 96 - 102

Tajfel H (1972) Experiments in a vacuum In J Isreal and H Tajfel (eds) *The Context of Social Psychology: A Critical Assessement*, Academic Press, London

Tajfel H (1979) Individuals and groups in social psychology *British Journal of Clinical and Social Psychology*, 18, 183 - 190

Tajfel H (1981) Social stereotypes and social groups In J C Turner & H Giles (eds) *Intergroup Behaviour*, Basil Blackwell, Oxford, UK

Tajfel H & Wilkes A L (1963) Classification and quantitative judgement *British Journal of Psychology*, 54, 101 - 114

Tajfel H, Billig M, Bundy R & Flament C (1971) Social categorization and intergroup behaviour *European Journal of Social Psychology*, 1, 149 - 178

Tata P R & Gudjonsson G H (unpublished) The effects of 'stress' and 'feedback' on interrogative suggestibility.

Taylor S E, Fiske S T, Etcoff N L & Ruderman A J (1978) Categorical and contextual bases of person memory and stereotyping *Journal of Personality and Social Psychology*, 36(7), 778 - 793

Terry R L & Snyder W G (1972) Veridicality of interpersonal perception based upon physiognomic cues *Journal of Psychology*, 81, 205 - 208

Terry R L (1975) Additional evidence for veridicality of perceptions based on physiognomic cues *Perceptual and Motor Skills*, 40, 780 - 782

Testa R J, Kinder B N & Ironson G (1987) Heterosexual bias in the perception of loving relationships of gay males and lesbians *Journal of Sex Research*, 23(2), 163 - 172

Thomson D M (1981) Person identification: Influencing the outcome *Australia and New Zealand Journal of Criminology*, 12, 49- 54

Thomson D M & Davies G M (1988) Introduction In G M Davies & D M Thomson (eds.) *Memory in Context: Context in Memory* John Wiley & Sons, Chichester

Thomson D M, Robertson S C & Vogt R (1982) Person recognition: The effect of context *Human Learning*, 1, 137 - 154

Thornton G R (1943) The effect upon judgements of personality traits of varying a single factor in a photograph *Journal of Social Psychology*, 28, 127 - 148

Toglia M P, Schlechter T M & Chevalier D S (1992) Memory for directly and indirectly experienced events *Applied Cognitive Psychology*, 6, 293 - 306

Tollestrup P A, Turtle J W & Yuille J C (in press) Actual victims and witnesses to robbery and fraud: An archival analysis In D F Ross, J D Read & M P Toglia (eds.): *Adult Eyewitness Testimony, Current Trends & Developments*, Cambridge University Press, New York

Treadway M & McCloskey M (1989) Effects of racial stereotypes on eyewitness performance: Implications of the real and the rumoured Allport and Postman studies *Applied Cognitive Psychology*, 3, 53 - 63

Tully B & Cahill D (1984) *Police interviewing of the mentally handicapped* The Police Foundation, London.

Tulving E (1983) *Elements of Episodic Memory* Oxford University Press, Oxford

Tulving E & Thompson D M (1973) Encoding specificity and retrieval processes in episodic memory *Psychological Review*, 80(5), 352 - 73

Turner J C (1975) Social comparison and social identity: Some prospects for intergroup behaviour *European Journal of Social Psychology*, 5, 5 - 34

Tversky A & Kahneman D (1973) Availability: A heuristic for judging frequency and probability *Cognitive Pychology*, 5, 207 - 232

Undeutsch U (1989) The development of statement reality analysis In J C Yuille (ed.) *Credibility Assessment* Kluwer Academic Publishers, London

Valentine T & Endo M (in press) Towards an exemplar model of face processing: The effects of race and distinctiveness *The Quarterly Journal of Experimental Psychology*

Valentine T (1988) Upside-down faces: A review of the effect of inversion upon face recognition *British Journal of Psychology*, 79, 471 - 491

Valentine T (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition *The Quarterly Journal of Experimental Psychology*, 43A(2), 161 - 204

Vana P (1986) The classroom modification of children's gender stereotyping of careers *Dissertation Abstracts International* 1547-A

Vivian S E & Gudjonsson G H (1986) Denial and memory performance in two cases of homicide *Medicine, Science & Law*, 26(1), 72 - 77.

Volgyesi F A (1966) *Hypnosis of Man and Animals* Baillière Tinsdale and Cassell, London

Wagenaar W A & Visser J (1979) The weather forecast under the weather *Ergonomics*, 22, 909 - 917

Walker-Smith G J, Gale A G and Findlay J M (1977) Eye movement strategies involved in face perception *Perception*, 6, 313 - 336

Wallach M A, Kogan N & Bem D J (1962) Group influence on individual risk taking *Journal of Abnormal and Social Psychology*, 65(2), 75 - 86

Waltman J L (1983) Nonverbal communication in interrogation: Some applications *Journal of Police Science and Administration* 166 - 169

Wegener H (1989) The present state of statement analysis In J C Yuille (ed.) *Credibility Assessment* Kluwer Academic Publishers, London

Weitzenhoffer A M (1958) *Hypnotism: An objective study in suggestibility* John Wiley & Sons: New York.

Wells G L (1978) Applied eyewitness testimony research: System variables and estimator varaibles *Journal of Personality and Social Psychology*, 36, 1546 - 1557

Wells G L (in press) The perceived credibility of eyewitnesses In M D MacLeod & C N Macrae (eds.) *Stereotypes in the Criminal Justice System* Falconer Press: UK

Wells G L, Leippe M R & Ostrom T M (1979) Guidelines for empirically assessing the fairness of a lineup *Law and Human Behavior*, 3, 285 - 293

Wells G L & Loftus E F (1979) *Eyewitness Testimony* Cambridge University Press: New York

Wells M (1989) *Guidelines for investigative interviewing of child victims of sexual abuse* The Metropolitan Chairman's Special Committee on Child Abuse Toronto Canada

Wilder D A (1978) Perceiving persons as a group: Effects on attribution of causality and beliefs *Social Psychology*, 36, 1361 - 1374

Wilder D A (1981) Perceiving persons as a group In D L Hamilton (ed) *Cognitive Processes in Stereotyping and Intergroup Behavior* Lawrence Erlbaum Associates: Hillsdale, New Jersey

Wilder D A & Allen V L (1978) Group membership and preference for information about other persons *Personality and Social Psychology Bulletin*, 4, 106 - 110

Williams G (1979) The authentication of statements to the police *Criminal Law Review*, 6 - 23

Williams L (1975) Applications of signal-detection parameters in a test of eyewitnesses to a crime report No. CR-20, Center for Responsive Psychology, New York

Williams M D & Hollan (1981) The process of retrieval from very long-term memory *Cognitive Science*, 5, 87 - 119

Wilson P (1968) Perceptual distortion of height as a function of ascribed academic status, *Journal of Social Psychology*, 74, 97 - 102

Wiseman R (1991) Differential recall of a pseudo-psychic demonstration as a function of belief in the paranormal University of Edinburgh Paper presented at the Parapsychological Association Convention

Woocher F D (1977) Did your eyes deceive you? Expert psychological testimony on the unreliability of eyewitness identification *Stanford Law Review*, 29, 969 - 1030

Woodhead M M, Baddeley A D & Simmonds D C V (1979) On training people to recognise faces *Ergonomics*, 22, 333 - 343

Word C O, Zanna M P & Cooper J (1974) The nonverbal mediation of self-fulfilling prophesies in interracial interaction *Journal of Experimental and Social Psychology*, 10, 109 - 120

Wright D B (1992) Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations Paper presented at Third European Conference of Law & Psychology, 16 - 19 September 1992

Wyer R S & Gordon S E (1984) The cognitive representation of social information In R S Wyer and T K Srull (eds.) *The Handbook of Social Cognition* Lawrence Erlbaum Associates: New Jersey.

Wyer R S, Bodenhausen G V & Srull T K (1984) The cognitive representation of persons and groups and its effect on recall and recognition memory *Journal of Experimental Social Psychology*, 20, 445 - 469

Yarbus A L (1967) *Eye Movements and Vision* (Translated by B Haigh) Plenum Press, New York

Young A W & Ellis H D (1989) *Handbook of Research on Face Processing* Elsevier: Amsterdam, Netherlands

Yuille J C (1984) Research and teaching with the police: A Canadian example *International Review of Applied Psychology*, 33 (1), 5 - 22

Yuille J C (1985) Paper presented to the World Conference on Police Psychology, December, Washington, United States

Yuille J C (1988) The systematic assessment of children's testimony *Canadian Psychology*, 29(3), 247 - 262

Yuille J C & Cutshall J L (1986) A case study of eyewitness memory of a crime *Journal of Applied Psychology*, 71(2), 291 - 301

Yuille J C & Kim C K (1987) A field study of the forensic use of hypnosis *Canadian Journal of Behavioural Science*, 19(4), 418 - 429

Yuille J C & McEwan N H (1985) The use of hypnosis as an aid to eyewitness testimony *Journal of Applied Psychology*, 70(2), 389 - 400

Zaragoza M S, McCloskey M & Jamis M (1987) Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13(1), 36 - 44

APPENDIX 1 - RATINGS OF BEHAVIOUR OF ACTORS IN VIDEO CLIPS, EXPERIMENT 1

TABLE 1 - RATED SENSITIVITY X BODY-SIZE PAIRING X PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	1.4248	0.2564	2.2708
AB	1/	28	54.2373	0.0000	100.0000
SZ AB	3/	28	2.4181	0.0873	4.4583
SZ SS	28				1.5938
SZ SS AB	28				1.8438

TABLE 2 - RATED MORALITY x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.5135	0.6763	0.7083
AB	1/	28	10.0674	0.0036	36.0000
SZ AB	3/	28	0.0816	0.9695	0.2917
SZ SS	28				1.3795
SZ SS AB	28				3.5759

TABLE 3 - RATED HOSTILITY x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.1483	0.9299	0.2708
AB	1/	28	0.2238	0.6398	0.5625
SZ AB	3/	28	2.1303	0.1188	5.3542
SZ SS	28				1.8259
SZ SS AB	28				2.5134

TABLE 4 - RATED COMPETITIVENESS x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.1652	0.1989	0.5573
AB	1/	28	27.9884	0.0000	112.8906
SZ AB	3/	28	0.3861	0.7638	1.5573
SZ SS	28				3.3728
SZ SS AB	28				4.0335

TABLE 5 - RATED INDIVIDUALITY x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	1.3879	0.2670	4.5573
AB	1/	28	0.0852	0.7726	0.3906
SZ AB	3/	28	2.3015	0.0988	10.5573
SZ SS	28				3.2835
SZ SS AB	28				4.5871

TABLE 6 - RATED LIBERALITY x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.6341	0.5993	1.8542
AB	1/	28	0.4715	0.4980	2.2500
SZ AB	3/	28	0.8469	0.4799	4.0417
SZ SS	28				2.9241
SZ SS AB	28				4.7723

TABLE 7 - RATED FAIRNESS x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.7763	0.5171	1.7917
AB	1/	28	5.7057	0.0239	20.2500
SZ AB	3/	28	1.1623	0.3416	4.1250
SZ SS	28				2.3080
SZ SS AB	28				3.5491

TABLE 8 - RATED INTELLIGENCE x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.5264	0.6677	0.8073
AB	1/	28	0.8384	0.3677	2.6406
SZ AB	3/	28	1.5528	0.2229	4.8906
SZ SS	28				1.5335
SZ SS AB	28				3.1496

TABLE 9 - RATED AGGRESSIVENESS x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.1699	0.9158	0.3542
AB	1/	28	0.0000	1.0000	0.0000
SZ AB	3/	28	1.9938	0.1378	8.6250
SZ SS	28				2.0848
SZ SS AB	28				4.3259

TABLE 10 - RATED LIKEABILITY x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	0.0806	0.9700	0.1823
AB	1/	28	9.5628	0.0045	34.5156
SZ AB	3/	28	0.7316	0.5419	2.6406
SZ SS	28				2.2612
SZ SS AB	28				3.6094

TABLE 11 - RATED COMPETENCE x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	2.7117	0.0639	3.8073
AB	1/	28	1.3247	0.2595	2.6406
SZ AB	3/	28	2.3490	0.0939	4.6823
SZ SS	28				1.4040
SZ SS AB	28				1.9933

TABLE 12 - RATED PRONENESS TO CRITICISM x BODY-SIZE PAIRING x PROTAGONIST/VICTIM

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	28	2.3710	0.0917	3.9323
AB	1/	28	9.9955	0.0038	34.5156
SZ AB	3/	28	0.0287	0.9933	0.0990
SZ SS	28				1.6585
SZ SS AB	28				3.4531

SZ - body-size combinations
AB - protagonist/victim
SS - subjects

APPENDIX 2 - INTENSITY RATINGS x BODY-SIZE OF ACTORS, EXPERIMENT 1

TABLE 1 - RATED INTENSITY x PERCEIVED BODY-SIZE PAIRING

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	1.2660	0.2939	3.7504
IN	3/	111	6.6465	0.0004	16.0221
SZ IN	6/	111	1.1187	0.3561	2.6969
SZ SS	37				2.9624
SZ SS IN	111				2.4106

TABLE 2 - RATED INTENSITY x ACTUAL BODY-SIZE PAIRING

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	3/	36	1.5223	0.2253	4.6229
IN	3/	108	7.4078	0.0001	18.1063
SZ IN	9/	108	0.3617	0.9507	0.8840
SZ SS	36				3.0368
SZ SS IN	108				2.4442

SZ - body-size combinations
IN - rated behaviour intensity
SS - subjects

APPENDIX 3 - ATTRIBUTION DATA x PERCEIVED BODY-SIZE OF ACTORS, EXPERIMENT 1

TABLE 1 - MAIN ANALYSIS OF VARIANCE, PERCEIVED SIZE x ATTRIBUTIONS GIVEN

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	0.0942	0.9103	0.3937
AT	3/	111	1.8736	0.1382	4.8746
SZ AT	6/	111	1.0390	0.4040	2.7032
SZ SS	37/				4.1785
SZ SS AT	111/				

TABLE 2 - ATTRIBUTIONS TO EXTERNAL FORCES, PERCEIVED BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	0.1028	0.9026	0.4162
SZ SS	37				4.0479

TABLE 3 - ATTRIBUTIONS TO INTERNAL FORCES, PERCEIVED BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	1.417	0.2553	3.3584
SZ SS	37				2.3701

TABLE 4 - ATTRIBUTIONS TO SITUATIONAL FORCES, PERCEIVED BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	0.4266	0.6559	1.7359
SZ SS	37				4.0689

TABLE 5 - ATTRIBUTIONS TO "OTHER" FORCES, PERCEIVED BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SZ	2/	37	0.5243	0.5963	1.6587
SZ SS	37				3.1635

SZ - perceived body-size pairings
SS - subjects
AT - attributions given

APPENDIX 4 - ATTRIBUTION DATA x ACTUAL BODY-SIZE OF ACTORS, EXPEIRMENT 1

TABLE 1 - ATTRIBUTIONS TO EXTERNAL FORCES, ACTUAL BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SV	1/	18	0.0857	0.7730	0.2250
SH	1/	18	0.1427	0.7101	0.6250
SV SH	1/	18	0.4623	0.5052	2.0250
SV SS	18				2.6250
SV SS SH	18				4.3806

TABLE 2 - ATTRIBUTIONS TO INTERNAL FORCES, ACTUAL BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SV	1/	18	1.4015	0.2519	3.0250
SH	1/	18	0.2312	0.6364	0.6250
SV SH	1/	18	1.5632	0.2272	4.2250
SV SS	18				2.1583
SV SS SH	18				2.7028

TABLE 3 - ATTRIBUTIONS TO SITUATIONAL FORCES, ACTUAL BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SV	1/	18	0.5145	0.4824	2.0250
SH	1/	18	0.4706	0.5014	2.0250
SV SH	1/	18	0.0058	0.9401	0.0250
SV SS	18				3.9361
SV SS SH	18				4.3028

TABLE 4 - ATTRIBUTIONS TO "OTHER" FORCES, ACTUAL BODY-SIZE PAIRINGS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SV	1/	18	5.9024	0.0258	15.6250
SH	1/	18	0.4353	0.5177	1.2250
SV SH	1/	18	1.9990	0.1745	5.6250
SV SS	18				2.6472
SV SS SH	18				2.8139

SV - size of victim (large or small)
SH - size of protagonist (large or small)
SS - subjects

APPENDIX 5 - SCRIPT FOR VIDEO CLIP, EXPERIMENT 1

Dilemma of choice: An electrical engineer must choose between staying in his present job at a modest but adequate salary, or moving to a new job which offers better pay but no long-term security.

A: Reads out dilemma

B: He should definitely take it.

A: Why? His present job offers long-term security.

B: What long-term security? Electrical engineers are dependent upon how much people can afford to pay for electrical engineers. You're like a public servant. You're dependent on the economy as a whole.

A: But if he'd got a steady job that gives him a decent income why should he risk it? This guy obviously has a place in life where he's comfortable. Our society teaches peoplethat they have to have a lot of money. People are taught money equals succss.

B: Who says he's motivated by that? For a start it says it's only an adequate salary. He's probably having to push himself to get that.

A: It doesn't mean that at all. It says it's adequate. It's clear he should stay where he is. Everyone these days is looking for a get-rich-quick scheme, but look how many have fallen flat on their faces. People shouldn't be looking for big money the whole time.

B: But he's not going to make big money. He's not likely to strike it rich or anything.

A: Money is the only thing this job can offer him that he doesn't have already. If I were him I'd stay where I was.

B: Well if I was an electrical engineer I'd take it. I think you're being condescending just because he's an electrical engineer. You think it's not for electrical engineers to take risks.

A: You're just assuming that's what I think.

B: Do you think anyone should take risks to get more money.

A: Yes.

B: Then you think that the sort of people who should be electrical engineers aren't the sort of people who should take risks.

A: Don't make assumptions about what I think. You're not thinking about all the ramifications of his decision. You don't know if he has a family, or a mortgage to pay off, or whether he's in debt, or what commitments he has. You just think he should take risks regardless and not consider long-term security or the people who could be affected. Do you think my father likes scrimping and saving all the time?

B: Your father?

A: Yes, my father. An electrical engineer.

B: Well I didn't know. Has he ever been offered a better job with no long-term security?

A: Yes he has. Are you're saying he didn't take it because he's an unambitious electrical engineer. You don't know anything about what might motivate someone to turn it down, you're just saying that they're afraid of taking risks.

B: No, I'm just saying that with an uncertain economy you've got to take what you're offered. A better job with more pay is a better risk than a supposedly more secure job.

A: So, basically, my father is an idiot and can't weigh up the odds.

B: I'm not judging your father, I'm just -

A: You are you're telling me -

Both start shouting; B administers "ambiguous shove" to A.

APPENDIX 6 - ORIGINAL TARGET STIMULI - EXPERIMENT 2

TABLE 1 - ORIGINAL TARGET STIMULI, RATINGS OF INTELLIGENCE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
TG	5/	30	0.1377	0.9822	0.3778
TG SS	30				2.7444

TABLE 2 - ORIGINAL TARGET STIMULI, RATINGS OF ATTRACTIVENESS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
TG	5/	30	0.5547	0.7336	0.7611
TG SS	30				1.3722

TABLE 3 - ORIGINAL TARGET STIMULI, RATINGS OF AGGRESSIVENESS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
TG	5/	30	0.1013	0.9911	0.1778
TG SS	30				1.7556

TABLE 4 - ORIGINAL TARGET STIMULI, RATINGS OF HONESTY

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
TG	5/	30	0.1082	0.9897	0.2278
TG SS	30				2.1056

TG - original target faces
SS - subjects

APPENDIX 7 - LIKENESS OF REGENERATIONS TO ORIGINAL TARGET STIMULI - EXPERIMENT 2

TABLE 1 - EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	20	0.0323	0.8592	0.0600
BA	1/	20	4.3942	0.0510	8.1667
GP BA	1/	20	0.1516	0.7012	0.2817
GP BA SS	20				1.8585

TABLE 2 - EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	20	0.0167	0.8983	0.0417
BA	1/	20	0.0429	0.8381	0.1067
GP BA	1/	20	0.9671	0.3371	2.4067
GP BA SS	20				2.4885

GP - reconstruction condition
BA - stereotype presentation time
SS - subjects

APPENDIX 8 - MAIN ANALYSES OF VARIANCE - EXPERIMENT 2

TABLE 1 - EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	20	4.6889	0.0426	3.3602
BA	1/	20	1.5492	0.2276	1.1102
GP BA	1/	20	2.3028	0.1448	1.6502
TR	1/	20	0.1065	0.7476	0.1302
GP TR	1/	20	18.2131	0.0004	22.2769
BA TR	1/	20	2.4938	0.1300	3.0502
GP BA TR	1/	20	1.4105	0.2489	1.7252
GP BA SS	20				0.7166
GP BA SS TR	20				1.2231

TABLE 2 - EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	20	2.7045	0.1157	3.3602
BA	1/	20	0.0605	0.8082	0.0752
GP BA	1/	20	0.4710	0.5004	0.5852
TR	1/	20	12.2261	0.0023	11.7019
GP TR	1/	20	3.8503	0.0638	3.6852
BA TR	1/	20	0.2092	0.6523	0.2002
GP BA TR	1/	20	0.7577	0.3944	0.7252
GP BA SS	20				1.2425
GP BA SS TR	20				0.9571

APPENDIX 9 - EFFECT OF TIME OF STEREOTYPE PRESENTATION - EXPERIMENT 2

TABLE 1 - EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
BA	1/	46	0.0538	0.8175	0.0752
BA SS	46				1.3967

TABLE 2 - EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
BA	1/	46	0.0636	0.8673	0.0987
BA SS	46				1.5460

APPENDIX 10 - SINGLE TRAIT JUDGMENTS ON REGENERATED FACES - EXPERIMENT 2b

TABLE 1 - STEREOTYPE CONDITION x RATED ATTRACTIVENESS

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	22	6.1107	0.0216	7.0417
GP SS	22				1.1523

TABLE 2 - STEREOTYPE CONDITION x RATED INTELLIGENCE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	22	0.0041	0.9497	0.0038
GP SS	22				0.9194

APPENDIX 11 - BIOGRAPHIES PRESENTED, EXPERIMENTS 2 AND 3

In this experiment we are interested in the way in which people form impressions of other people. Please read the following short biography very carefully. It outlines briefly the life of the man in the accompanying picture. Your task here is to try to form as accurate a picture in your mind as you can about this character. For instance, try to decide what are his likes and dislikes; might he be an optimistic person or a pessimist, etc. You will be given about 3 minutes to complete this task.

The man pictured here, for present purposes simply called X, was born in Leamington Spa in England in 1965. He lived there with his parents until 1967, when the family moved to Stratford. His brother was born in Stratford in 1972, at about the same time that their father lost his job. The family became very poor. Both parents took a number of odd jobs, but the family continued to get deeper into debt. Meanwhile, X was taken out of the public school which he had started, and sent to a comprehensive.

Within three years, his father had found another job. The family paid off its debts, and life became easier. X started secondary school; which was where he discovered rock music, which has been a passion with him ever since.

In 1981, X's parents bought a house in the Scottish Highlands. X went to live with his grandparents while he studied for 'O' levels. Afterwards, he went to live in the Highlands for 2 years. He found life there very dull, and holidayed in England whenever possible.

X attended university in Edinburgh. In his first year there he met the girl whom he was later to marry. He had originally intended to to study Philosophy, but ended up with a 2i [or 2ii] degree in Psychology instead. After graduating, X managed to find himself a fairly well-paid office job in Edinburgh, at which he is still working.

OR

X attended university in Edinburgh. In his first year there he met the girl whom he was later to marry. He had originally intended to study Philosophy, but ended up with a degree in Psychology instead. After graduating, X managed to find himself a fairly well-paid office job in Edinburgh. Life was looking quite bright for X until the early part of the year, when he was arrested and charged with committing a theft [or assault].

APPENDIX 12 - ORIGINAL TARGET STIMULI - EXPERIMENT 3a AND b

TABLE 1 - TRAIT RATINGS OF 4 ORIGINAL TARGET STIMULI

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
TG	3/	16	0.3025	0.8232	0.4900
TR	4/	64	2.3262	0.0657	2.4600
TG TR	12/	64	0.3530	0.9747	0.3733
TG SS	16				1.6200
TG SS TR	64				1.0575

TG - original target bodies
TR - four character traits, + body-size ratings
SS - independent raters

TABLE 2 - ACTUAL BODY-SIZE RECONSTRUCTED x CONDITION, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	3.1934	0.0804	221.6139
GP SS	48				69.3968

TABLE 3 - ACTUAL BODY-SIZE RECONSTRUCTED x CONDITION, EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	0.0156	0.9011	1.6914
GP SS	48				108.2902

GP - reconstruction condition
SS - reconstrcuted images

APPENDIX 13 - LIKENESS OF REGENERATED BODIES TO ORIGINAL STIMULI - EXPERIMENT 3a AND b

TABLE 1 - LIKENESS TO ORIGINAL STIMULI, REGENERATED BODIES, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	44	8.9213	0.0046	9.3633
BA	1/	44	6.1487	0.0171	6.4533
GP BA	1/	44	1.8294	0.1831	1.9200
GP BA SS	44				1.0495

TABLE 2 - LIKENESS TO ORIGINAL STIMULI, REGENERATED BODIES, EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	44	0.3902	0.5354	0.6075
BA	1/	44	1.3923	0.2444	2.1675
GP BA	1/	44	0.5829	0.4492	0.9075
GP BA SS	44				1.5567

GP - reconstruction condition
BA - stereotype presentation time
SS - independent raters

APPENDIX 14 - INITIAL ANALYSES OF VARIANCE, ALL VARIABLES - EXPERIMENT 3a AND b

TABLE 1 - INITIAL ANALYSIS OF VARIANCE, EXPERIMENT a. RATINGS OF AGGRESSIVENESS, HONESTY AND BODY-SIZE x STEREOTYPE CONDITION x STEREOTYPE PRESENTATION TIME

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	44	2.2813	0.1381	2.5069
BA	1/	44	0.0063	0.9370	0.0069
GP BA	1/	44	0.2429	0.6246	0.2669
TR	2/	88	2.7052	0.0724	2.4803
GP TR	2/	88	0.9237	0.4008	0.8469
BA TR	2/	88	1.0128	0.3674	0.9286
GP BA TR	2/	88	1.6309	0.2016	1.4953
GP BA SS	44				1.0989
GP BA SS TR	88				0.9169

TABLE 2 - INITIAL ANALYSIS OF VARIANCE, EXPERIMENT b. RATINGS OF INTELLIGENCE AND ATTRACTIVENESS x STEREOTYPE CONDITION x STEREOTYPE PRESENTATION TIME

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	44	0.3199	0.5746	0.5104
BA	1/	44	0.3199	0.5746	0.5104
GP BA	1/	44	0.0943	0.7603	0.1504
TR	1/	44	19.9882	0.0001	8.2837
GP TR	1/	44	0.0493	0.3629	0.5500
BA TR	1/	44	0.0251	0.8748	0.0104
GP BA SS	44				1.5958
GP BA SS TR	44				0.4144

GP - stereotype condition
BA - stereotype presentation time
TR - character traits and body-size ratings
SS - independent raters

APPENDIX 15 - EFFECT OF TIME OF STEREOTYPE PRESENTATION ON JUDGMENTS OF REGENERATED BODIES - EXPERIMENT 3

TABLE 1 - EFFECT OF TIME OF STEREOTYPE PRESENTATION, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
BA	1/	142	0.0000	1.0000	0.0000
BA SS	142				1.0044

TABLE 2 - EFFECT OF TIME OF STEREOTYPE PRESENTATION, EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
BA	1/	94	0.4917	0.4849	0.5104
BA SS	94				1.0380

BA - time of stereotyype presentation
SS- independent judges

q

APPENDIX 16 - SINGLE TRAIT JUDGMENTS ON REGENERATED BODIES - EXPERIMENT 3a AND b

TABLE 1 - STEREOTYPE CONDITION x RATED AGGRESSIVENESS, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	2.3581	0.1315	3.6300
GP SS	46				1.5393

TABLE 2 - STEREOTYPE CONDITION x RATED HONESTY, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	0.0354	0.8516	0.0208
GP SS	46				0.5882

TABLE 3 - STEREOTYPE CONDITION x RATED BODY-SIZE, EXPERIMENT a

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	4.1053	0.0486	4.0833
GP SS	46				0.9946

TABLE 4 - STEREOTYPE CONDITION x RATED INTELLIGENCE, EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	0.2108	0.6483	0.1633
GP SS	46				0.7749

TABLE 5 - STEREOTYPE CONDITION x RATED ATTRACTIVENESS, EXPERIMENT b

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	46	1.5196	0.2239	1.5052
GP SS	46				0.9905

GP - stereotype condition
SS - independent judges

APPENDIX 17 - GENERATION OF ITEMS USED IN THE FORCED-CHOICE RECOGNITION TASKS

For each of the critical items used in their experiment, Bellezza and Bower generated four alternatives - a "less" and a "more" heterosexual and lesbian alternative. Along with their neutral items, all of the "less" heterosexual and lesbian alternatives were pilotted on 100 undergraduates, who were asked to indicate on a nine-point scale to what degree each statement was consistent with a "lesbian" or a "heterosexual" stereotype. Only if more than half of the subjects indicated that the alternative was consistent with a stereotype was that alternative retained. Listed below are all the alternatives pilotted on the subjects, along with the number of subjects who found that alternative consistent with which stereotype.

	lesbian	heterosexual
1 Betty's father		
b) was kind to her	0	20
c) was mean to her	28	0
2 As a child Betty liked movies that		
b) were comedies	0	12
c) were adventures	1	2
3 In her early teens Betty		
b) went to some parties where there were boys and girls	1	4
c) went to parties with her girlfriends	24	0
4 Physically, Betty was		
b) not unattractive	0	20
c) not particularly attractive	28	0
5 In high school Betty		
b) went out with boys occasionally	0	72
c) went out with a boy once	8	64
6 In high school Betty		
b) had a crush on Mr Griffin	1	72
c) felt very close to Ms Griffin	60	1
7 In college Betty		
b) lived in a coed dorm	1	50
c) lived in a girls' dorm	32	1
8 In college Betty		
b) was able to get along with her roommate	8	8
c) got along great with her roommate	12	0

9 In college Betty		
b) became friends with some women students who lived in her dorm	4	13
c) became close friends with some women students who lived in her dorm	4	8
10 In college Betty was able to reveal her most intimate secrets		
b) to some of her close female friends	3	11
c) only to her female friends	4	4
11 Early in college Betty went out places in a group		
b) with her female friends	0	36
c) only with her female friends	58	0
12 After Betty started working at the animal clinic		
b) Dr and Mrs Sawyer became second parents to her	0	16
c) Mrs Sawyer became a second parent to her	16	8
13 While working at the animal clinic Betty would tell		
b) Dr and Mrs Sawyer about her problems and anxieties	5	20
c) Mrs Sawyer about her problems and anxieties	14	4
14 While working at the animal clinic Betty		
b) came to like John very much	0	86
c) asked John's advice on a number of things	0	32
15 While working at the animal clinic Betty		
b) got along with Sue	12	8
c) liked Sue	14	0
16 While a member of a theatre group Betty		
b) admired Anne	37	0
c) admired Anne tremendously	31	0
17 In college Betty		
b) liked to dress up often	10	29
c) liked to dress up occasionally	14	0
18 In college Betty		
b) wore make up occasionally	5	28
c) rarely wore make up	8	0
19 While a member of a theatre group Betty		
b) thought that Robert was attractive	0	96
c) thought that Robert was interesting	5	43

20 While a member of a theatre group Betty		
b) hoped her relationship with Robert would continue	0	86
c) thought that her relationship with Robert wouldn't continue	0	56
21 In college Betty		
b) dated guys often	0	88
c) dated guys more often than in high school	0	100
22 On her dates in college Betty		
b) had some really good times	0	66
c) had a good time occasionally	16	32
23 For awhile at college Betty		
b) went steady with David	0	100
c) went out often with David	0	92
24 While at college Betty		
b) was seriously involved with David	0	100
c) felt that David was a good friend	0	34
25 While at college Betty		
b) reluctantly cut off her relationship with David	4	72
c) decided to stop seeing David	8	23
26 In college Betty felt that		
b) she could communicate with both men and women	4	13
c) she could communicate better with women	37	4
27 The name of Betty's high school guidance counselor was:		
a) Ms Pennington	12	4
b) Mr Griffin	0	5
28 At college Betty would often go with her close friends to		
a) art museums	0	8
b) bars	0	3
29 During college Betty was		
a) seemingly interested in medicine	0	0
b) not interested in medicine	0	0
30 Her guidance counselor advised Betty to		
a) wait as long as possible before deciding on a major	9	0
b) become a doctor	0	0

31 Betty and her family moved to the midwest when Betty was		
a) 8 years old	0	3
b) 12 years old	0	0
32 Betty's roommate at college was		
a) Joan	0	4
b) Jane	0	0
33 Betty's major in college was		
a) biochemistry	10	0
b) microbiology	8	0
34 Socially Betty		
a) would reveal personal information only to her close friends	3	0
b) open and freely revealed personal information to many people	3	0
35 Betty went to elementary school at		
a) Redwood school	0	1
b) Claremont school	0	1
36 Betty's best friend in high school was		
a) Patty	4	8
b) Peggy	3	8
37 The name of the university that Betty attended was		
a) Mullin University	0	0
b) Midwestern University	0	3
38 The name of Betty's high school was		
a) Lincoln High School	4	0
b) Midvale High School	9	0
39 Betty lived in the town of		
a) Midvale	0	0
b) Midfield	0	0

While all of the *neutral* items above were judged to be neutral by the subjects, very few critical items were judged to be consistent with the sexual stereotype in question. Therefore, for those items not judged to be consistent with the sexual stereotype, Bellezza and Bower's "more" heterosexual and lesbian alternatives were presented to the subjects. The results are detailed below.

	lesbian	heterosexual
1 Betty's father		
a) was kind to her and her mother	10	53
d) was mean to her and her mother	60	4

2 As a child Betty liked movies that		
a) were love stories	1	55
d) had women as heroes	70	4
3 In her early teens Betty		
a) went to many parties where there were boys and girls	6	12
d) went to parties only with her girl friends	36	7
4 Physically, Betty was		
a) attractive	7	78
d) rather unattractive	57	1
5 In high school Betty		
d) avoided going out with boys	88	0
7 In college Betty		
d) wanted to live in a girls' dorm	68	5
8 In college Betty		
a) considered having a male roommate	6	64
b) became very attached to her roommate	51	16
9 In college Betty		
a) became friends with male students who lived nearby	8	71
d) became intimately involved with some women who lived in her dorm	89	0
10 In college Betty was able to reveal her most intimate secrets		
a) to some of her close male and female friends	12	51
d) only to female friends she was very close to	62	8
11 Early in college Betty went out places in a group		
a) with her male and female friends	7	12

[NB a new alternative had to be invented to replace this: the new alternative read "only with her male friends"]

12 After Betty started working at the animal clinic		
a) Dr Sawyer became a second parent to her	2	66
d) only Mrs Sawyer became a second parent to her	73	8

13 While working at the animal clinic Betty would tell		
a) Dr Sawyer about her problems and anxieties	10	61
d) only Mrs Sawyer about her problems and anxieties	51	3
14 While working at the animal clinic Betty		
d) respected John's knowledge	5	7
15 While working at the animal clinic Betty		
a) liked John more than Sue	0	54
d) liked Sue very much	51	17
16 While a member of a theatre group Betty		
a) thought that Anne was self-centred	7	8
d) was physically attracted to Anne	90	0
17 In college Betty		
a) was a smart dresser	17	19
d) never liked to dress up	19	6
18 In college Betty		
a) often wore make up	3	21
d) didn't like to wear make up	33	17
19 While a member of a theatre group Betty		
d) thought that Robert was uninteresting	13	12
20 While a member of a theatre group Betty		
d) was relieved that her relationship with Robert ended.	6	7
21 In college Betty		
d) never dated guys	76	1
22 On her datesin college Betty		
d) never had a good time	53	1
23 For awhile at college Betty		
a) wanted to go out only with David	0	44
d) went out occasionally with David	2	40
24 While at college Betty		
d) couldn't communicate with David	23	12
25 While at college Betty		
d) got tired of David	17	5
26 In college Betty felt that		
a) she could communicate only with men	5	53
b) she could communicate only with women	67	0

14 of the 17 critical iterms were thus found usable - ie. 14 of them had a pair of alternatives, one of which was judged consistent with a lesbian stereotype and one with a heterosexual stereotype. These were the items used in the forced-choice recognition tasks.

The standard test procedure tested alternatives rated as *lesbian* against alternatives rated as *heterosexual* in every case. The modified test procedure tested the *correct* alternative in every case against an alternative that had been rated *neutral*.

The resultant test forms are detailed below:

STANDARD TEST

1 Betty's father
  a) was kind to her and her mother
  b) was mean to her and her mother

2 Betty and her family moved to the Midwest when Betty was
  a) 8 years old
  b) 12 years old

3 Betty lived in the town of
  a) Midvale
  b) Midfield

4 As a child, Betty
  a) liked movies that were love stories
  b) liked movies that had women as heroes

5 Betty went to elementary school at
  a) Redwood school
  b) Claremont school

6 Betty's best friend in high school was
  a) Patty
  b) Peggy

7 Physically, Betty was
  a) not particularly attractive
  b) not unattractive

8 In high school Betty
  a) went out with boys occasionally
  b) avoided going out with boys

9 In high school Betty
  a) had a crush on Mr Griffin
  b) felt very close to Ms Griffin

10 The name of Betty's high school was
  a) Lincoln High School
  b) Midvale High School

11 The name of Betty's high school guidance counselor was
  a) Mr Griffin
  b) Ms Pennington

12 Her guidance counselor advised Betty to
  a) wait as long as possible before deciding on a major
  b) become a doctor

13 The name of the university that Betty attended was
  a) Midwestern University
  b) Mullin University

14 In college Betty
  a) wanted to live in a coed dorm
  b) wanted to live in a girls' dorm

15 In college Betty
  a) considered having a male roommate
  b) became very attached to her roommate

16 Betty's roommate at college was
  a) Jane
  b) Joan

17 In college Betty
  a) became friends with some male students who lived nearby
  b) became intimately involved with some women who lived nearby

18 While in college, Betty went out in a group
  a) only with her male friends
  b) only with her female friends

19 At college Betty would often go with her close friends to
  a) art museums
  b) bars

20 After Betty started working at the animal clinic
  a) only Dr Sawyer became a second parent to her
  b) only Mrs Sawyer became a second parent to her

21 While working at the animal clinic Betty
  a) liked Sue more than John
  b) liked John more than Sue

22 In college Betty
  a) never dated guys
  b) dated guys often

23 On her dates in college Betty
  a) had some really good times
  b) never had a good time

24 In her decision whether to marry or pursue a career Betty
  a) thought that she wanted to become an actress
  b) thought that she wanted to become a doctor

25 In college Betty felt that
  a) she could communicate only with men
  b) she could communicate only with women

26 Betty's major in college was
  a) microbiology
  b) biochemistry

MODIFIED TEST

1 Betty's father
  a) was a kind man
  b) was mean to her and her mother

2 Betty and her family moved to the Midwest when Betty was
  a) 8 years old
  b) 12 years old

3 Betty lived in the town of
  a) Midvale
  b) Midfield

4 As a child, Betty
  a) liked movies that were love stories
  b) liked movies that were adventures

5 Betty went to elementary school at
  a) Redwood school
  b) Claremont school

6 Betty's best friend in high school was
  a) Patty
  b) Peggy

7 Physically, Betty was
  a) not particularly attractive
  b) neither particularly attractive nor particularly unattractive

8 In high school Betty
  a) rarely went out with anyone
  b) avoided going out with boys

9 In high school Betty
  a) admired Mr Griffin
  b) felt very close to Ms Griffin

10 The name of Betty's high school was
  a) Lincoln High School
  b) Midvale High School

11 The name of Betty's high school guidance counselor was
  a) Mr Griffin
  b) Ms Pennington

12 Her guidance counselor advised Betty to
  a) wait as long as possible before deciding on a major
  b) become a doctor

13 The name of the university that Betty attended was
  a) Midwestern University
  b) Mullin University

14 In college Betty
  a) lived in a girls' dorm
  b) wanted to live in a girls' dorm

15 In college Betty
  a) was able to get along with her roommate
  b) became very attached to her roommate

16 Betty's roommate at college was
  a) Jane
  b) Joan

17 In college Betty
  a) was friendly with some men and women who lived nearby
  b) became intimately involved with some women who lived nearby

18 While in college, Betty went out in a group
  a) only with her male and female friends
  b) only with her female friends

19 At college Betty would often go with her close friends to
  a) art museums
  b) bars

20 After Betty started working at the animal clinic
  a) only Dr Sawyer became a second parent to her
  b) Dr and Mrs Sawyer became second parents to her

21 While working at the animal clinic Betty
  a) disliked John
  b) liked John more than Sue

22 In college Betty
  a) rarely dated guys
  b) dated guys often

23 On her dates in college Betty
  a) had a good time occasionally
  b) never had a good time

24 In her decision whether to marry or pursue a career Betty
  a) thought that she wanted to become an actress
  b) thought that she wanted to become a doctor

25 In college Betty felt that
  a) she could communicate with both men and women
  b) she could communicate only with women

26 Betty's major in college was
  a) microbiology
  b) biochemistry

APPENDIX 18 - MATERIALS USED, EXPERIMENTS 4 & 5

Biographies presented to the subjects

In this experiment we are interested in the way individuals form impressions of other people. You will read the story of one individual's life. Names and places have been changed to conceal the person's identity. What we want you to do is read the story of the person's life and think about what the person is like. Try to form an impression of the person. For example, what are her likes and dislikes, personality characteristics, wants and needs, etc.? Later, we will ask you questions about your reactions to this person. Your task is simply to develop as complete an understanding as you can of this person.

You will have 5 minutes to read this story.

CASE HISTORY

Betty K grew up in an urban environment on the West Coast of America. She had a stormy childhood since her parents were fighting most of the time. Her father, an engineer, was an abusive man who was mean to her and her mother. Finally, one day when Betty was 8 years old, her mother took her and her baby brother from their West Coast home and moved to the midwest to get away from their father.

After that, Betty's life became more tranquil. She and he mother and brother lived happily in a suburban home in the city of Midvale. Her mother worked as a legal secretary and made enough money so that Betty's family was fairly well off financially. Betty liked going to movies as a child, and particularly liked love stories. At Redwood School, where she went for most of her elementary education, her favourite teacher was Miss Brock, who gave Betty much encouragement and had a great impact on Betty's high motivation to do well academically.

Betty made lots of friends, and in her early teens began to attend some parties. At this time she also became particularly close to one girl named Patty. The two spent a lot of time together throughout junior high and high school. Although most of their girl friends began dating when they were 16 or 17, Betty and Patty, who were physically not very attractive, went out on dates only ocasionally. When they did it was usually together on double dates.

At Lincoln High School, Betty greatly admired her sophomore art teacher, Ms Griffin. She never told anyone about this, not even Patty. For some time Betty felt rejected because Ms Griffin showed no special interest in her. After about 2 months, however, she got over her disappointment.

Betty, who was always near the top of the class academically, decided to go to college after graduation on the advice of her guidance counsellor, Ms Pennington. She had thought of becoming a doctor, but hadn't really decided. Her guidance counsellor advised Betty to wait as long as possible before deciding on a major and to take a wide variety of courses at college to see exactly where her interests lay.

When she entered college at Mullin University, a school with about 15 000 students, she decided to live in a girls' dorm. She became very attached to her roommate, Joan. She also became friends with some male students who lived nearby, and they would all frequently hold long rap sessions late into the night. During these sessions, Betty would often find herself revealing some of her most intimate secrets, and others would do the same. She really enjoyed the intimacy she felt with these close friends. She often went out with her male friends and they did things as a group, for instance attending art museums and plays.

During her first year in college, Betty got a part-time job working for Dr Sawyer, a vetinarian. Betty loved working with animals and caring for them. This was a maturing experinece for her because she had to take on important responsibilities in the running of the clinic and in dealing with the pets that were brought in, and their owners.

Also, Betty got to know Dr and Mrs Sawyer very well. Dr Sawyer became like a second parent to her. She would sometimes take a break from her work and tell Dr Sawyer about her many problems and anxieties. Two other college students also worked for the Sawyers - John and Sue. Betty got on particularly well with John, and sought his advice on many things. He was like the older brother that she never had. She also liked Sue, and they became friends. They had a great many things in common.

About this time, Betty joined an amateur theatre group and became acquainted with many of its members. However, the acquaintances never seemed to become more than that. One of the members, Anne, was a beautiful and talented actress. Betty admired her tremendously and occasionally tried to dress and make herself up to look like Anne. However, she was never able to become one of her close friends. Robert, another member of the group, had coffee with Betty a few times. Betty thought that Robert was attractive, but doubted that their relationship would continue. In fact, their relationship didn't seem to go anywhere.

Betty started dating guys more often than she did in high school. She dated guys almost every weekend, although she never had a really good time on these dates. She went out often with one guy named David, a business major, for about six months. She became seriously involved with David, but he was much more serious about the relationship that she was, and wanted to get married. Somewhat reluctantly, she decided to break up with David. She had reached a point where she couldn't really see her relationship with any man as being more than friendship, while with her female friends she could be herself. She could communicate more easily with women. She did not have to make an effort to impress them, or win their approval, as she did with men.

Betty had a hard time deciding on a major. Joan, her roommate, seemed quite engrossed in her major, psychology, but Betty was equally interested in so many things, she didn't know which one to pick. She finally decided on biochemistry, since if she wanted to go to medical school it would probably help her. Also, she did very well in her biochemistry classes.

By the time she reached her junior year of college, Betty had to decide what to do with her life. Late in her junior year, she decided to devote her full time to a medical career, assuming her grades were good enough to get into medical school. Being a woman of high ability and motivation, she is likely to be highly succesful in her chosen career.

Stereotyping paragraph presented to subjects

We would like you to know several other facts about Betty's life:
During her senior year in college, Betty met a man/lesbian who introduced her to sexual/homsexual activity. Betty felt exhileated that she had finally found herself. She [married this man] and went on to a successful medical career living with her husband/lesbian lover. She found life in general very satisfying.

Questions to increase salience of stereotype label

Please answer the following questions briefly in the spaces provided:

When you were reading Betty's history, did you believe that she would eventually adopt a homosexual lifestyle?

How many specific instances can you remember form Betty K's history that indicated that she would eventually adopt a homosexual lifestyle?

What percentage of women with Betty's type of background do you think are able to adopt a lifestyle that they find satisfying?

List 3 events in Betty's life that you believe were the most important in forming her later values and attitudes.

APPENDIX 19 - SUMMARY ANOVA TABLES FOR RECOGNITION CONDITIONS - EXPERIMENT 4

TABLE 1 - MAIN ANALYSIS OF VARIANCE FOR RECOGNITION TASKS x STEREOTYPE LABEL x INFORMATION GENERATED IN RESPONSE TO CRITICAL ITEMS

SOURCE	DF1/	DF2	F	PValue	MEAN SQUARE
GP	3/	116	1.5280	0.2110	0.8806
IN	3/	348	1010.8050	0.0000	976.1917
GP IN	9/	348	3.7785	0.0001	3.6491
GP SS	116/				0.5763
GP SS IN	348/				0.9658

GP - test procedure
IN - information generated
SS - subjects

APPENDIX 20 - EFFECT OF STEREOTYPE LABELS - EXPERIMENT 4

TABLE 1 - ANALYSIS OF VARIANCE FOR HETEROSEXUAL LABEL, STANDARD vs MODIFIED TEST PROCEDURE x CORRECT INFORMATION x LESBIAN AND HETEROSEXUAL

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	1.7617	0.1896	2.1333
IN	1/	58	2.0142	0.1612	1.6333
GP IN	1/	58	4.1106	0.0472	3.3333
GP SS	58/				1.2109
GP SS IN	58/				0.8109

TABLE 2 - ANALYSIS OF VARIANCE FOR LESBIAN LABEL, STANDARD vs MODIFIED TEST PROCEDURE x CORRECT INFORMATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	1.2406	0.2699	2.1333
IN	1/	58	6.5034	0.0134	6.5333
GP IN	1/	58	1.1945	0.2789	1.2000
GP SS	58/				1.7195
GP SS IN	58/				1.0046

GP - test procedure
IN - information generated
SS - subjects

APPENDIX 21 - ERRORS MADE IN RECOGNITION TASK CONDITIONS - EXPERIMENT 4

TABLE 1 - STANDARD vs MODIFIED TEST PROCEDURE, HETEROSEXUAL LABEL, FALSE ALARMS MADE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	28.9239	0.0000	12.6750
IN	1/	58	3.0880	0.0841	1.8750
GP IN	1/	58	2.3195	0.1332	1.4083

TABLE 2 - STANDARD vs MODIFIED TEST PROCEDURE, LESBIAN LABEL, FALSE ALARMS MADE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	13.2571	0.0006	8.5333
IN	1/	58	0.1871	0.6669	0.1333
GP IN	1/	58	0.7484	0.3906	0.5333

TABLE 3 - STANDARD vs MODIFIED TEST PROCEDURE, HETEROSEXUAL LABEL, FALSE ALARMS MADE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	28.9239	0.0000	12.6750
IN	1/	58	3.0880	0.0841	1.8750
GP IN	1/	58	2.3195	0.1332	1.4083

GP - test procedure
IN - information produced

APPENDIX 22 - NEUTRAL INFORMATION PRODUCED - EXPERIMENT 4

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	3/	116	0.5567	0.6447	1.6556
GP IN	116/				2.9741

GP - test procedure
IN - neutral information generated

APPENDIX 23 - FREE RECALL RESULTS - EXPERIMENT 4

TABLE 1 - CORRECT INFORMATION GENERATED x STEREOTYPE CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	5.2390	0.0257	7.0042
IN	3/	174	165.4276	0.0000	183.5153
GP IN	3/	174	6.8146	0.0002	7.5597

TABLE 2 - FALSE ALARMS x STEREOTYPE LABEL CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	3.9835	0.0507	0.8333
IN	1/	58	4.2647	0.0434	0.8333
GP IN	1/	58	4.2647	0.0434	0.8333
GP SS	58/				0.2692
GP SS IN	58/				0.1954

TABLE 3 LESBIAN FALSE ALARMS x STEREOTYPE LABEL CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	5.3506	0.0243	1.6667
GP IN	58				0.3115

TABLE 4- NEUTRAL HITS x STEREOTYPE LABEL CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	58	1.1416	0.2897	10.4167

GP - stereotype label
IN - false alarms generated

APPENDIX 24 - SUMMARY ANOVA TABLES, EXPERIMENT 5

TABLE 1 - INITIAL ANALYSIS OF VARIANCE, ALL VARIABLES

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	3/	76	8.1220	0.0001	4.7585
IN	3/	228	201.6994	0.0000	316.8583
GP IN	9/	228	13.9513	0.0000	21.9167

TABLE 2 - COMPARISON OF HETEROSEXUAL vs. LESBIAN LABEL

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	78	5.1058	0.0266	3.6125
IN	3/	234	138.3138	0.0000	316.8583
GP IN	3/	234	2.8178	0.0398	6.4542

TABLE 3 - COMPARISON OF IMMEDIATE vs. DELAYED TESTING

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	78	9.8831	0.0024	6.6125
IN	3/	234	188.7778	0.0000	316.8583
GP IN	3/	234	32.3037	0.0000	54.2208

TABLE 4 - NEUTRAL INFORMATION GENERATED vs. LABEL RECEIVED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	78	5.6945	0.0194	10.0000
IN	1/	78	8.7021	0.0042	140.6250
GP IN	1/	78	1.0458	0.3096	16.9000

TABLE 5 - NEUTRAL INFORMATION GENERATED vs. DELAY CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	78	5.9605	0.0169	10.5062
IN	1/	78	14.4850	0.0003	138.7562
GP IN	1/	78	56.0119	0.0000	536.5563

GP - testing conditions
IN - information generated

TABLE 6 - CORRECT INFORMATION GENERATED - HETEROSEXUAL LABEL

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IN	1/	39	49.8370	0.0000	86.1125
SS	39				2.4484
SS IN	39				1.7279

SS - subjects

TABLE 7 - CORRECT INFORMATION GENERATED - HOMOSEXUAL LABEL

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IN	1/	39	30.8857	0.0000	52.8125
SS	39				3.2048
SS IN	39				1.7099

TABLE 8 - INCORRECT RESPONSES x LABEL

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
GP	1/	78	1.0181	0.3161	1.8063
IN	1/	78	6.7569	0.0112	9.5063
GP IN	1/	78	1.6037	0.2091	2.2562
GP SS	78				1.7742
GP SS IN	78				

APPENDIX 25 - DESCRIPTION OF STAGED INCIDENT

Mid-way through a first year social psychology lecture, the double doors at the back of the lecture theatre (i.e., behind the undergraduates) were opened, and a girl entered the theatre. She was 21 years old. Her hair was in an old-fashioned bun. She was wearing glasses, grey coat, red shirt, black basque, black lace gloves, black mini-skirt, fishnet stockings with black and red garter, black shoes. She was also wearing very large and ornate gold-coloured earrings and a large quantity of make-up. She was carrying a portable stereo playing *New York, New York*; which alerted the undergraduates to her presence. This music was chosen not for its instant recognisability. Once she had the attention of the audience, the girl called out the lecturer's name. She then ostentatiously removed her glasses, placed them in the pocket of her coat and proceeded to dance down the seventeen steps from the back of the lecture theatre to the point at which the lecturer was standing. On the way, she removed coat and gloves and dropped them on the stairs, and unbuttoned her shirt. At the front of the lecture theatre, she removed her garter and handed it to the lecturer with the words "Happy retirement from Athene Gordon". As she was leaning forward to kiss the lecturer, the right hand front door of the lecture theatre (in front of the undergraduates) opened, and a young man entered, shouting "Mary Margaret!" He was wearing a brightly-coloured, red and blue ski jacket, blue and white T-shirt, blue trousers, and cowboy boots. The man marched across to where the girl was standing (this involved walking almost the entire width of the lecture theatre, thus exposing him for the longest possible period to the undergraduates). On the way, he shouted "I thought you weren't going to do this any more", to which the girl replied, "But it's for a good cause". The man said "What about your Finals? What about your Modern History Finals?" to which the girl did not reply. Removing his jacket, the man said "Look at yourself!" He covered the girl with the jacket and told her "I've been waiting half an hour for you." He then took her with him out of the right-hand front door of the lecture theatre (the same door through which he had entered), the couple arguing all the way. As they passed the benches at the front of the lecture theatre, the girl's hip caught the trestle-table that had been placed there, knocking a large plastic tub and two pens to the floor.

APPENDIX 26 - EFFECT OF TIME AND STEREOTYPIC INFORMATION ON MEMORY FOR STAGED INCIDENT - EXPERIMENT 6

TABLE 1 - MAJOR ANALYSIS OF VARIANCE, EFFECT OF TIME AND STEREOTYPIC INFORMATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	0.0825	0.7778	0.6454
PH	2/	30	3.4759	0.0439	19.4994
ST PH	2/	30	0.4001	0.6738	2.2445
ST SS	15				7.8184
ST SS PH	30				5.6098

TABLE 2 - EFFECT OF TIME AND STEREOTYPIC INFORMATION, CONSISTENT INFORMATION RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	2.7574	0.1176	0.7003
PH	2/	30	2.3791	0.1099	0.3731
ST PH	2/	30	2.8793	0.0718	0.4515
ST SS	15				0.2540
ST SS PH	30				0.1568

TABLE 3 - EFFECT OF TIME AND STEREOTYPIC INFORMATION, INCONSISTENT INFORMATION RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	0.8460	0.3722	0.2868
PH	2/	30	0.8071	0.4556	0.4866
ST PH	2/	30	0.1566	0.8558	0.0944
ST SS	15				0.3390
ST SS PH	30				0.6029

TABLE 4 - EFFECT OF TIME AND STEREOTYPIC INFORMATION, INCIDENTAL INFORMATION RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	0.9148	0.3540	0.0473
PH	2/	30	1.0012	0.3794	0.0613
ST PH	2/	30	0.3612	0.6998	0.0221
ST SS	15				0.0517
ST SS PH	30				0.0613

TABLE 5 - EFFECT OF TIME, GROUP 1

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
PH	2/	10	1.4286	0.2846	0.8889
SS	5				0.8889
SS PH	10				0.6222

TABLE 6 - EFFECT OF TIME, GROUP 2

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
PH	2/	12	1.2000	0.3349	1.0000
SS	6				0.7619
SS PH	12				0.8333

ST - presentation or otherwise of stereotypic information
PH - phase of experiment
SS - subjects

APPENDIX 27 - WITHIN-CONDITION EFFECTS - EXPERIMENT 6

TABLE 1 - STEREOTYPE CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
PH	2/	18	0.1765	0.8397	0.0238
CO	1/	18	64.8000	0.0000	7.7143
PH CO	2/	18	0.6000	0.5594	0.0714
PH SS	18				0.1349
PH SS CO	18				0.1190

TABLE 2 - NON-STEREOTYPE CONDITION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
PH	2/	27	1.9884	0.1565	0.3167
CO	1/	27	10.2857	0.0034	2.4000
PH CO	2/	27	2.7857	0.0795	0.6500
PH SS	27				0.1593
PH SS CO	27				0.2333

PH - phase of experiment
CO - two types of information
SS - subjects

APPENDIX 28 - ERRORS MADE - EXPERIMENT 6

TABLE 1 - ERRORS MADE, CONSISTENT INFORMATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	0.1498	0.7041	0.0101
PH	2/	30	0.9696	0.3908	0.0810
ST PH	2/	30	0.0302	0.9703	0.0025
ST SS	15				0.0673
ST SS PH	30				0.0835

TABLE 2 - ERRORS MADE, INCONSISTENT INFORMATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	3.0612	0.1006	1.0768
PH	2/	30	0.1336	0.8755	0.0213
ST PH	2/	30	1.6100	0.2167	0.2566
ST SS	15				0.3517
ST SS PH	30				0.1594

TABLE 3 - ERRORS MADE, INCIDENTAL INFORMATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	2.4706	0.1368	5.4902
PH	2/	30	3.9819	0.0293	5.8490
ST PH	2/	30	0.6180	0.5457	0.9078
ST SS	15				2.2222
ST SS PH	30				1.4689

ST - presentation or otherwise of stereotypic information
PH - phase of experiment
SS - subjects

APPENDIX 29 - ACCURACY OF RECALL - EXPERIMENT 6

TABLE 1 - TOTAL INFORMATION RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	45	3.8040	0.0574	365.4185
PH	2/	45	4.2754	0.0200	423.0709
ST PH	2/	45	0.6106	0.5475	60.4230
ST PH SS	45				98.9540

TABLE 2 - INCIDENTAL INFORMATION RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	20	1.0679	0.3138	912.1622
PH	2/	20	0.7515	0.4845	641.8919
ST PH	2/	20	0.7515	0.4845	641.8919
ST PH SS	20				854.1667

ST - presentation or otherwise of stereotypic information
PH - phase of experiment
SS - subjects

APPENDIX 30 - MEMORY FOR CLOTHING - EXPERIMENT 6

TABLE 1 - RECALL OF KISSOGRAM'S CLOTHING x PHASE x STEREOTYPE PRESENTATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	0.4823	0.4923	0.5179
PH	2/	30	0.4164	0.6632	0.5773
ST PH	2/	30	0.7276	0.4914	1.0087
ST SS	15				1.0740
ST SS PH	30				1.3863

TABLE 2 - RECALL OF BOYFRIEND'S CLOTHING x PHASE x STEREOTYPE PRESENTATION

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
ST	1/	15	7.1460	0.0174	3.3280
PH	2/	30	2.5365	0.0960	2.2490
ST PH	2/	30	1.7404	0.1927	1.5431
ST SS	15				0.4657
ST SS PH	30				0.8867

TABLE 3 - STYLE vs COLOUR - KISSOGRAM'S CLOTHING

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SY	2/	48	0.2401	0.7875	0.5588
SC	1/	48	9.3603	0.0036	2.8333
SY SC	2/	48	0.2267	0.7980	0.0686
SY SS	48				2.3272
SY SS SC	48				0.3027

TABLE 4 - STYLE vs COLOUR - BOYFRIEND'S CLOTHING

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
SY	2/	48	2.4849	0.0940	1.9216
SC	1/	48	96.0831	0.0000	45.3333
SY SC	2/	48	1.0805	0.3475	0.5098
SY SS	48				0.7733
SY SS SC	48				0.4718

ST - presentation or otherwise of stereotypic information
PH - phase of experiment
SS - subjects
SY - style of clothing
SC - colour of clothing

APPENDIX 32 - INITIAL ANALYSIS OF VARIANCE - EXPERIMENT 7

TABLE 1 - ANALYSIS OF VARIANCE, INTERVIEW GROUP x INTERVIEW PHASE x INFORMATION TYPE RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	2.1229	0.1323	3.4704
PH	2/	84	152.5074	0.0000	233.7370
IG PH	4/	84	3.0098	0.0226	4.6130
IN	5/	210	87.0676	0.0000	100.2459
IG IN	10/	210	2.1701	0.0208	2.4985
PH IN	10/	420	21.3993	0.0000	25.9652
IG PH IN	20/	420	1.6511	0.0386	2.0033
IG SS	42				1.6347
IG SS PH	84				1.5326
IG SS IN	210				1.1514
IG SS PH IN	420				1.2134

IG - interview group
PH - interview phase
IN - information type
SS - subjects

APPENDIX 33 - FINAL TWO PHASES - EXPERIMENT 7

TABLE 1 - ANALYSIS OF VARIANCE, INTERVIEW GROUP x FINAL TWO PHASES x INFORMATION TYPE RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	0.3457	0.7098	0.2796
PH	1/	42	140.2299	0.0000	100.5352
IG PH	2/	42	0.1886	0.8288	0.1352
IN	5/	210	55.9806	0.0000	37.3263
IG IN	10/	210	0.8893	0.5441	0.5930
PH IN	5/	210	31.0355	0.0000	22.0730
IG PH IN	10/	210	1.1056	0.3592	0.7863
IG SS	42				0.8090
IG SS PH	42				0.7169
IG SS IN	210				0.6668
IG SS PH IN	210				0.7112

APPENDIX 34 - UNIVARIATE ANALYSES OF VARIANCE - EXPERIMENT 7

TABLE 1 - VOICE-OVER ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	0.2564	0.7750	0.2741
PH	2/	84	66.7651	0.0000	64.3630
IG PH	4/	84	1.2870	0.2817	1.2407
IG SS	42				1.0688
IG SS PH	84				0.9640

TABLE 2 - CONVERSATIONAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	2.9636	0.0625	9.4741
PH	2/	84	56.1394	0.0000	146.1407
IG PH	4/	84	2.4699	0.0508	6.4296
IG SS	42				3.1968
IG SS PH	84				2.6032

TABLE 3 - VIDEO ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	2.5455	0.0905	0.2667
PH	2/	84	4.9000	0.0097	0.6222
IG PH	4/	84	1.4875	0.2133	0.1889
IG SS	42				0.1048
IG SS PH	84				0.1270

TABLE 4 - CONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	0.8260	0.4448	0.7185
PH	2/	84	65.6967	0.0000	60.2741
IG PH	4/	84	2.2849	0.0669	2.0963
IG SS	42				0.8698
IG SS PH	84				0.9175

TABLE 5 - INCONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	5.4772	0.0077	1.6519
PH	2/	84	8.8926	0.0003	2.5407
IG PH	4/	84	1.9704	0.1065	0.5630
IG SS	42				0.3016
IG SS PH	84				0.2857

TABLE 6 - VISUAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	1.0153	0.3710	2.6741
PH	2/	84	43.3375	0.0000	103.2074
IG PH	4/	84	0.8616	0.4906	2.0519
IG SS	42				2.6339
IG SS PH	84				2.3815

APPENDIX 35 - MAIN ANALYSIS OF VARIANCE - ERRORS - EXPERIMENT 7

TABLE 1 - ANALYSIS OF VARIANCE, ERRORS, INTERVIEW GROUP x INTERVIEW PHASE x INFORMATION TYPE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	1.2466	0.2979	0.3593
IN	5/	210	24.5594	0.0000	5.6170
IG IN	10/	210	3.4233	0.0004	0.7830
PH	2/	84	0.9519	0.3901	0.2111
IG PH	4/	84	2.3213	0.0634	0.5148
IN PH	10/	420	2.1998	0.0170	0.4793
IG IN PH	20/	420	1.8190	0.0170	0.3963
IG SS	42				0.2882
IG SS IN	210				0.2287
IG SS PH	84				0.2218
IG SS IN PH	420				0.2179

APPENDIX 36 - FINAL TWO PHASES - EXPERIMENT 7

TABLE 1 - ERRORS MADE x INTERVIEW GROUP x INTERVIEW PHASE x INFORMATION TYPE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	4.6148	0.0154	1.3722
IN	5/	210	18.5061	0.0000	3.9989
IG IN	10/	210	2.2573	0.0159	0.4878
PH	1/	42	10.2327	0.0026	2.2685
IG PH	2/	42	1.5119	0.2322	0.3352
IN PH	5/	210	2.5681	0.0280	0.4552
IG IN PH	10/	210	0.9884	0.4546	0.1752
IG SS	42				0.2724
IG SS IN	210				0.2161
IG SS PH	84				0.2217
IG SS IN PH	420				0.1772

APPENDIX 37 - UNIVARIATE ANALYSES OF VARIANCE - EXPERIMENT 7

TABLE 1 - VOICE-OVER ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	0.4108	0.6658	0.1852
PH	2/	84	4.5822	0.0129	1.6074
IG PH	4/	84	0.9397	0.4452	0.3296
IG SS	42				0.4508
IG SS PH	84				0.3508

TABLE 2 - CONVERSATIONAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	4.1774	0.0221	1.0963
PH	2/	84	8.1807	0.0006	2.8741
IG PH	4/	84	4.3223	0.0031	1.5185
IG SS	42				0.2624
IG SS PH	84				0.3513

TABLE 3 - VIDEO ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	1.2250	0.3040	0.0519
PH	2/	84	1.3659	0.2608	0.0296
IG PH	4/	84	1.3659	0.2527	0.0296
IG SS	42				0.0423
IG SS PH	84				0.0217

TABLE 4 - CONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	1.6517	0.2039	0.1556
PH	2/	84	2.9822	0.0561	0.2667
IG PH	4/	84	1.7396	0.1489	0.1556
IG SS	42				0.0942
IG SS PH	84				0.0894

TABLE 5 - INCONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	0.5698	0.5700	0.0519
PH	2/	84	8.1395	0.0006	0.7407
IG PH	4/	84	0.5698	0.6853	0.0519
IG SS	42				0.0910
IG SS PH	84				0.0910

TABLE 6 - VISUAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	42	4.1408	0.0228	1.8667
PH	2/	84	13.7222	0.0000	5.4889
IG PH	4/	84	1.7222	0.1527	0.6889
IG SS	42				0.4508
IG SS PH	84				0.4000

APPENDIX 38 - ACCURACY DATA - EXPERIMENT 7

TABLE 1 - ACCURACY OF RECALL x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	113	1.3949	0.2521	404.9179
PH	2/	113	6.7977	0.0016	1973.2446
IG PH	4/	113	0.6444	0.6320	187.0540
IG PH SS	113				290.2825

APPENDIX 39 - INITIAL ANALYSIS OF VARIANCE - EXPERIMENT 8

TABLE 1 - ANALYSIS OF VARIANCE, INTERVIEW GROUP x INTERVIEW PHASE x INFORMATION TYPE RECALLED

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.7709	0.1894	2.2296
PH	2/	54	177.5376	0.0000	137.3907
IG PH	4/	54	0.2106	0.9315	0.1630
IN	5/	135	59.6439	0.0000	51.7085
IG IN	10/	135	2.5154	0.0083	2.1807
PH IN	10/	270	16.2166	0.0000	12.8819
IG PH IN	20/	270	0.8486	0.6529	0.6741
IG SS	27				1.2591
IG SS PH	54				0.7739
IG SS IN	135				0.8670
IG SS PH IN	270				0.7944

IG - interview group
PH - interview phase
IN - information type
SS - subjects

APPENDIX 40 - UNIVARIATE ANALYSES OF VARIANCE - EXPERIMENT 8

TABLE 1 - VOICE-OVER ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	6.5763	0.0047	4.3111
PH	2/	54	22.4236	0.0000	26.0778
IG PH	4/	54	0.5111	0.7278	0.5944
IG SS	27				0.6556
IG SS PH	54				1.1630

TABLE 2 - CONVERSATIONAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	2.0704	0.1457	5.7333
PH	2/	54	44.4394	0.0000	64.3000
IG PH	4/	54	0.6796	0.6091	0.9833
IG SS	27				2.7691
IG SS PH	54				1.4469

TABLE 3 - VIDEO ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	0.3364	0.7173	0.0444
PH	2/	54	4.6901	0.0132	0.4111
IG PH	4/	54	1.2676	0.2941	0.1111
IG SS	27				0.1321
IG SS PH	54				0.0877

TABLE 4 - CONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.0481	0.3644	0.7000
PH	2/	54	106.4028	0.0000	46.6333
IG PH	4/	54	0.9887	0.4216	0.4333
IG SS	27				0.6679
IG SS PH	54				0.4383

TABLE 5 - INCONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.0556	0.3619	0.2111
PH	2/	54	4.0851	0.0223	0.7111
IG PH	4/	54	0.7340	0.5728	0.1278
IG SS	42				0.2000
IG SS PH	84				0.1741

TABLE 6 - VISUAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.8982	0.1693	2.2333
PH	2/	54	47.2083	0.0000	66.0333
IG PH	4/	54	0.9056	0.4673	1.2667
IG SS	27				1.1765
IG SS PH	54				1.3988

APPENDIX 41 - ERRORS MADE - EXPERIMENT 8

TABLE 1 - ANALYSIS OF VARIANCE, ERRORS, INTERVIEW GROUP x INTERVIEW PHASE x INFORMATION TYPE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	9.2840	0.0009	2.9056
PH	2/	54	14.5889	0.0000	3.0889
IG PH	4/	54	2.2959	0.0709	0.4861
IN	5/	135	10.6058	0.0000	2.1722
IG IN	10/	135	1.4051	0.1844	0.2878
IN PH	10/	270	3.3915	0.0003	0.4844
IG IN PH	20/	270	1.5285	0.0713	0.2183
IG SS	27				0.3130
IG SS PH	54				0.2117
IG SS IN	135				0.2048
IG SS IN PH	270				0.1428

APPENDIX 42 - UNIVARIATE ANALYSIS, ERRORS MADE - EXPERIMENT 8

TABLE 1 - VOICE-OVER ONLY INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	8.2552	0.0016	1.4778
PH	2/	54	12.2324	0.0000	2.1444
IG PH	4/	54	5.1021	0.0015	0.8944
IG SS	27				0.1790
IG SS PH	54				0.1753

TABLE 2 - CONVERSATIONAL INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.4372	0.2552	1.1444
PH	2/	54	3.7907	0.0288	1.8111
IG PH	4/	54	0.3023	0.8751	0.1444
IG SS	27				0.7963
IG SS PH	54				0.4778

TABLE 3 - CONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	7.7294	0.0022	0.8111
PH	2/	54	6.0330	0.0043	0.6778
IG PH	4/	54	2.7692	0.0363	0.3111
IG SS	27				0.1049
IG SS PH	54				0.1123

TABLE 5 - INCONSISTENT INFORMATION x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	27	1.0000	0.3811	0.0111
PH	2/	54	1.0000	0.3746	0.0111
IG PH	4/	54	1.0000	0.4157	0.0111
IG SS	42				0.0111
IG SS PH	84				0.0111

APPENDIX 43 - ACCURACY DATA - EXPERIMENT 8

TABLE 1 - ACCURACY OF RECALL x INTERVIEW GROUP x PHASE

SOURCE	DF1/	DF2	F	P VALUE	MEAN SQUARE
IG	2/	70	24.6202	0.0000	7129.2797
PH	2/	70	5.6067	0.0055	1623.5378
IG PH	4/	70	4.9995	0.0013	1447.7151
IG PH SS	70				289.5705